DOCUMENT RESUME

ED 192 667 HE 013 137

AUTHOR Lawrence, Judith K.; Green, Kenneth C.
TITLE A Question of Quality: The Higher Education Ratings Game. AAHE-ERIC/Higher Education Research Report No. 5, 1980.
INSTITUTION American Association for Higher Education, Washington, D.C.; ERIC Clearinghouse on Higher Education, Washington, D.C.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 80
CONTRACT 400-77-0073
NOTE 76p.
AVAILABLE FROM American Association for Higher Education, One Dupont Circle, Suite 780, Washington, DC 20036 ($3.00 members, $4.00 nonmembers)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Academic Standards; Accreditation (Institutions); Educational Assessment; *Educational Quality; Educational Research; *Evaluation Criteria; Graduate Study; *Higher Education; *Institutional Evaluation; Professional Education; Program Evaluation; Reputation; *Research Methodology; State Standards; Undergraduate Study

ABSTRACT

Studies concerning attributes of quality in higher education as defined in academic studies are reviewed. Separate reviews are presented for studies of quality at the graduate level, in professional programs, and at the undergraduate level. Academe's continuing attempts to quantify "quality" so as to measure it empirically rather than subjectively through reputational ratings are examined. In addition, accreditation and state program review, both of which exemplify external approaches to assessing quality in American higher education, are discussed. In academic studies, usually conducted by researchers from the higher education community, assessments have focused on identifying the best institutions (or graduate departments). It is concluded that whether based on peer review or on the application of a set of traditionally-used quantifiable indicators, such assessments ignore about 99 percent of the nation's higher education institutions. It is suggested that these rankings serve to reinforce the hierarchical structure of the system, whereby material and human wealth tend to be concentrated in a few institutions. It is also noted that the teaching-learning function of higher education has been virtually ignored in quality assessments. Conclusions and recommendations as to how quality in higher education might be better defined and how methods of assessing quality might be improved are presented. A bibliography is included. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED192667

AAHE-ERIC/Higher Education Research Report No. 5, 1980

A Question of Quality: The Higher Education Ratings Game

Judith K. Lawrence and Kenneth C. Green

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Prepared by

ERIC®
Clearinghouse on Higher Education
The George Washington University

Published by

AAHE
American Association for Higher Education

HE 013137

ERIC Clearinghouse on Higher Education
One Dupont Circle, Suite 630
Washington, D.C. 20036

American Association for Higher Education
One Dupont Circle Suite 780
Washington, D.C. 20036

This publication was prepared with funding from the National Institute of Education, U.S. Department of Education, under contract no. 400-77-0073. The opinions expressed in this report do not necessarily reflect the positions or policies of NIE or the Department.

Foreword

The issue of quality in higher education has been around for a long time. New institutions have sought to imitate the learning formats and organizational structures of "quality" institutions with the expectation that, by doing this, they also would acquire a similar reputation for quality. But many of these aspiring institutions fail to achieve their goal because they do not fully comprehend that quality is based on more than just format and structure.

With the prediction of enrollment declines and increased financial pressures caused by inflation, the issue of quality has taken on even greater significance for every institution, even those of high repute. Identifying, nuturing, and promoting the distinctive qualities of an institution may mean the difference between survival and extinction. This means it is now crucial that every institution develop a more sophisticated understanding of what quality connotes and how it can be measured.

In this report, written by Judith K. Lawrence and Kenneth C. Green of the Higher Education Research Institute, Los Angeles, California, the question of quality is examined from the perspective of how quality has been measured in the past. The authors first review studies that analyze the reputation of graduate education and professional programs. From this examination quantifiable indicators of quality are identified and reviewed in light of undergraduate education. The authors conclude their report with a discussion of quality in relation to accreditation and state program review.

For those who are concerned with identifying and measuring quality, this report is the first step to understanding how quality has been defined in the past. With this foundation the reader will be better able to examine and measure quality indicators at his or her own institution.

Jonathan D. Fife
Director
ERIC Clearinghouse on Higher Education

Acknowledgements

This work was supported by a grant from the Exxon Education Foundation to the Higher Education Research Institute as part of an ongoing study of quality in American higher education. To our colleagues at the Institute we extend appreciation for their support and interest. Thanks go particularly to Lewis C. Solmon for his helpful comments, Laura Kent for her editorial assistance, and Molly Landi for her skillful and patient preparation of the manuscript.

We dedicate this work to the memory of Allan M. Cartter, a founding officer of the Institute, whose contribution to the assessment of quality in American higher education remains a milestone.

Judith K. Lawrence
Kenneth C. Green
Los Angeles, April 1980

Contents

Overview 1

Reputational Studies of Graduate Education 4
Ranking Studies 4
Commentaries and Critiques 8
Recent Quality Assessments 12
The Master's Degree 14
Summary and Conclusions 15

Assessments of Professional Program Quality 17
Studies by the Academic Community 18
Studies by the Professions 21
Summary and Conclusions 22

Quantifiable Indicators of Quality 23
Scholarly Productivity as an Indicator of Quality 25
Other Quantifiable Indicators of Quality 27
Summary and Conclusions 31

Quality Assessment at the Undergraduate Level 32
Traditional Academic Studies 32
An Example of a Popular College Guide 36
The Input-Environment-Outcome Model 38
Summary and Conclusions 41

Other Dimensions and Concerns in Quality Assessment 42
Accreditation 43
State Program Review 47
Summary and Conclusions 51

Conclusions 53

Bibliography 56

Overview

> Quality . . . you know what it is, yet you don't know what it is. But that's self-contradictory. But some things *are* better than others, that is, they have more quality. But when you try to say what the quality is, apart from the things that have it, it all goes *poof!* There's nothing to talk about. But if you can't say what Quality is, how do you know what it is, or how do you know that it even exists? If no one knows what it is, then for all practical purposes it doesn't exist at all. But for all practical purposes it really *does* exist. What else are grades based on? Why else would people pay fortunes for some things and throw others in the trash pile? Obviously some things are better than others . . . but what's the 'betterness'?. . . So round and round you go, spinning mental wheels and nowhere finding anyplace to get traction. What the hell is Quality? What *is* it? (Pirsig 1974, p. 184)

Anyone concerned with quality in American higher education is caught in the quandry described by Pirsig: "What the hell is Quality? What is it?" We know that a degree earned at Harvard is different (this word, rather than "better," is used advisedly) from one earned at Ohio State, is different from one earned at Pratt Institute, is different from one earned at Antioch College, is different from one earned at Oral Roberts University, and so on, ad infinitum.

And certainly we know intuitively, if no other way, that there is a tremendous range in program and institutional quality among the 3,000-plus colleges and universities in the U.S. higher education system. Yet assessments of quality by particular persons, for particular purposes, in particular contexts, result in a variety of quality attributes that leave the meaning of the term elusive.

Turning to the literature, this monograph deals primarily with the attributes of quality defined in academic studies, separately reviewing studies of quality at the graduate level (chapter 1), in professional programs (chapter 2), and at the undergraduate level (chapter 4), and separately addressing academe's continuing attempts to quantify "quality" so as to measure it empirically (chapter 3) rather than subjectively through reputational ratings. In addition, chapter 5 discusses accreditation and state program review, both of which exemplify external approaches to assessing quality in American higher education. The

final chapter presents conclusions and recommendations concerning quality assessment in higher education.

Starting with Hughes' pioneering rating study in 1925 through those conducted under the prestigious sponsorship of the American Council on Education (Cartter 1966; Roose and Andersen 1970), reputational ratings of graduate education have formed the foundation of quality assessments of academe by academics. Such ratings are based on peer review, wherein programs are rated by faculty panels in the same discipline, as experts, and their results reflect the prominence of graduate education and of faculty research in the system. To the detriment of the undergraduate level, of the teaching-learning function, and of the diversity that characterizes the nation's higher education system, 50 years of reputational ratings have consistently identified 20 or 30 outstanding institutions, leaving them to vie with each other for the highest absolute rank in the hierarchial structure, and virtually ignoring the rest of our colleges and universities.

In the domain of the top institutions—whether graduate or professional or undergraduate—an enormous range of material and human resources have been shown to correlate with reputational prestige and with each other. Foremost among the material resources are institutional size, library size, research-related variables such as funds available, and faculty salary. Foremost among the human resources are faculty and student abilities, background, and achievements—particularly scholarly productivity in lending visibility to individual scholars and their current institutions. In other words, material and human wealth tend to be concentrated in a few institutions.

Yet this description is surely inadequate with respect to the meaning and measurement of quality in the nation's pluralistic higher education system. Indeed, the researchers have been increasingly mindful of errors of omission, errors inherent in prioritizing the graduate level, the top domain of institutions, and so forth. Moreover, they have become increasingly cognizant of the myriad differences that exist in higher education—from domain to domain, from level to level, and from discipline to discipline—especially regarding goals and objectives. Thus, they continue to investigate how the perplexing concept of quality might be broadened to accommodate the strengths of such differences. Clearly aware of having confused quantity with quality, whether through conducting reputational studies or studies based on quantifiable indicators or combining both, recent quality assessments show consensus on a number of needs:

1. If comparisons must be made, they should be made between similar types of institutions, at the same level, in the same disciplines, and so forth.

2. Quality assessments must identify program goals and objectives and be referenced to them.

3. Quality assessments must be based on a variety of attributes.

4. The meaning of "quality" is—and should be—as varied as the purposes behind an assessment, the measurement criteria used, and the group or groups conducting the assessment; herein lie the value and limitations of quality assessments.

5. The teaching-learning function of higher education has been virtually ignored in quality assessments. Conceptually and methodologically, the value-added, input-environment-output model merits further investigation.

The quality question will be a major concern in the coming decade. Quality is inextricably tied to such issues as equality of access and choice, post-baccalaureate employment and the value of a college degree, curriculum structure, and student development and outcomes. Only by understanding how quality has been assessed can we know how and in what contexts it should be measured and which interventions will yield improvements.

Reputational Studies of Graduate Education

The best-known quality studies in American higher education are those sponsored by the American Council on Education (ACE): Allan Cartter's *An Assessment of Quality in Graduate Education* (1966) and its replication, *A Rating of Graduate Programs* (1970) by Kenneth Roose and Charles Andersen. These studies, which rank graduate programs by institution, have served as prototypes for quality assessments of graduate education and as catalysts for examining quantifiable, allegedly objective indicators of quality. As a result of their influence, the literature on quality in higher education informs us most about the graduate level, about reputational ratings, and about the correlations between reputational ratings and quantifiable indicators (Hartnett, Clark, and Baird 1978).

Ranking studies

When Raymond Hughes (1925) conducted his pioneering reputational study of graduate programs in 1924, only 65 universities in the U.S. awarded the doctoral degree. His study ranked 38 universities in 20 graduate disciplines according to the number of top scholars they employed, as listed by panels of scholars from each field. By 1934 the number of institutions awarding the doctorate had increased to 106; a second Hughes study (1934) rated 59 universities in 35 fields as "adequate" or "distinguished," according to faculty raters' assessments of staff and facilities for the preparation of doctoral candidates. The two Hughes studies went well beyond their stated purpose of informing undergraduates about graduate programs, and established important precedents for quality ratings: a focus on the graduate rather than the undergraduate level, reliance on the opinions of academicians themselves rather than of outside observers, the assigning of numerical positions (ranks) to institutions on the basis of this "informed opinion," and an emphasis on the nation's leading institutions.

More than two decades passed before any attempt was made to update Hughes' work. The first cross-discipline post-war study, conducted by Keniston in 1957, ranked 24 graduate programs at 25 institutions as part of a comparative self-study at the University of Pennsylvania (Keniston 1959). From a list of the 25 institutions leading in doctorate production, the raters—

department chairmen selected from the institutional members of the American Association of Universities—were asked to name the top five departments in their fields on a combined measure of doctoral program quality and faculty quality. Keniston then compared his aggregated, rank-ordered list of the top 20 institutions with Hughes' results to show "what changes have taken place in the course of a generation" (p. 116).

The major weaknesses of the Hughes and Keniston studies are summarized in the introduction to the 1966 ACE ratings. According to Cartter, these earlier rankings reflect geographical and rater biases, though in Hughes' case geographical bias may have been inevitable since at that time the most distinguished universities were concentrated in the Northeast and the Midwest. Cartter comments on three other flaws in the Keniston study: failure to separate measures of faculty quality from measures of educational quality, failure to anticipate that raters would overrank their alma maters, and the choice of department chairmen as raters. On this last point, Cartter asserts that chairmen are not necessarily the most distinguished scholars in their fields; that they are not typical of their peers in age and rank, specialization, or knowledge of the academic scene; and that because they tend to be older and more conservative, and thus to favor those institutions which have traditionally produced the largest number of doctorates, their ratings reflect outdated perceptions.

These criticisms guided the design of the ACE studies; great care was taken to achieve an equitable geographical distribution and to assure the representativeness of institutions and raters. Cartter defends the validity of using peer raters to evaluate graduate education:

> The present study is a survey of *informed* opinion. The opinions we have sought are what in a court of law would be called "the testimony of expert witnesses"—those persons in each field who are well qualified to judge, who by training are both knowledgeable and dispassionate, who through professional activities are competent to assess professional standards, and who by their scholarly participation within their chosen fields have earned the respect of their colleagues and peers (Cartter 1966, p. 8).

As with previous reputational studies, the informative value of subjective ratings of quality is assumed by the Cartter report, which has three stated purposes. The first is to update the Hughes and Keniston studies. The second purpose, directed to-

ward future policy, is "to widen the assessment to include all major universities in the United States, on the assumption that major expansion will not come from the 10-15 traditionally distinguished universities" (p. 3). The third purpose is to examine and compare subjective and objective measurement techniques.

Cartter surveyed the 100 institutions that, in 1961, formed the Council of Graduate Schools in the U.S., as well as the six universities that had granted "100 or more doctorates (spread over three or more fields) in the preceding decade" (p. 12). In all, 1,663 graduate programs in 29 disciplines were rated. The 4,000 survey respondents included department chairmen, distinguished senior scholars, and knowledgeable junior scholars who had completed their formal training not more than ten years earlier. From an alphabetized list of the 106 institutions, respondents were asked to rate each doctoral program in their own field of study on two components: quality of graduate faculty and effectiveness of the doctoral program. The response alternatives for rating faculty quality were "distinguished," "strong," "good," "adequate," "marginal," and "not sufficient"; respondents were told to limit the number of "distinguished" ratings to five. Response alternatives for the second component, program effectiveness, were "extremely attractive," "attractive," "acceptable," and "not attractive." An "insufficient information" option was included for both components. To determine the representativeness of the raters, the questionnaire also requested basic biographical information. The leading departments are ranked separately on faculty quality and program effectiveness, although in most fields the two lists are very similar. Whenever possible, the report presents Hughes' 1925 rankings and Keniston's 1959 rankings for comparison.

Considerable attention is given to the statistical validity and reliability of the ratings. Detailed analyses are presented for three fields: economics, to represent the social sciences; English, to represent the humanities; and physics, to represent the natural sciences. Replicating Hughes' methodology of using small panels of experts to rate these three fields, Cartter determined that a minimum of 50 knowledgeable persons are required for reliability.

Prior to Cartter's study, Somit and Tanenhaus (1964) had surveyed 830 political science departments on the basis of peer ratings and quantifiable indicators. Therefore, in assessing the validity of the reputational peer-rating approach used in the ACE study, Cartter chose political science as a fourth field. He reports a high correlation between the two sets of ratings of po-

litical science departments and calls the Somit and Tanenhaus study "a preview of the ACE survey" (Cartter 1966, p. 100).

Cartter also compared peer ratings with so-called objective measures. For instance, he found that the rankings produced by subjective responses were consistent with the institutional rankings found by Bowker (1964), who used the enrollment of graduate award recipients in institutional programs as a criterion. Other quantifiable indicators used by Cartter for comparison with peer ratings are faculty salary, library resources, and publication indexes; in each case, the results tend to corroborate the study's findings. Cartter thus concludes:

> It seems likely that if one were to include enough factors in constructing a so-called objective index--allowing for variations in institutional size and a university's commitments to certain fields of study--the results of our subjective assessment would be almost exactly duplicated (Cartter 1966, p. 118).

Cartter was unwilling to aggregate departmental ratings to produce institutional rankings for three reasons. First, not all institutions offer doctorates in every field. Second, it would be very difficult to assign weights to various fields. Third, departmental specialization is the chief organizing principle in academe.

The academic community's response to the Cartter report was overwhelming. The report stimulated widespread comment and critique, and by 1970 more than 26,000 copies had been distributed. The 1970 Roose-Andersen study essentially replicates the 1966 ACE study, fulfilling Cartter's commitment to do a five-year follow-up study lest reputations become "writ in stone" (Cartter 1966, p. 8). Roose and Andersen (1970) assert that the purpose of their report is informational; it is not intended "to inflate or deflate institutional egos. It is hoped that readers will think in terms of quality ranges rather than specific pecking orders" (p. 33).

Thus, the Roose-Andersen report presents ranges of scores rather than "absolute" raw departmental scores. In addition to this change, the word "quality" is omitted from the title. The authors state:

> Since it is evident . . . that the appraisal is of faculty and programs as reflected by their reputations rather than as they partake of specific components of an amorphous attribute called "quality," we have resolved to use as a title simply a description of the book's contents, *A Rating of Graduate Programs*. . . (Roose and Andersen 1970, p. xi).

The follow-up study extended the sample from 29 to 36 fields and from 106 to 131 institutions. A total of 2,626 graduate programs are thus surveyed, representing a 58 percent increase over the number of programs rated in the 1966 report.

Although both ACE reports explicitly and implicitly discourage the aggregation of departmental data into institutional "scores," other researchers ("American Education Council . . .," 1971; Magoun 1966; Morgan, Kearney, and Regens 1976; National Science Foundation 1969; Petrowski, Brown, and Duffy 1973) were quick to make such aggregations, perhaps because comparing programs and institutions with one another is "an almost inevitable byproduct of the American competitive spirit" (Clark 1976, p. 35). Moreover, researchers and government officials find total institutional scores useful in providing a developmental view of higher education and in facilitating national planning (Magoun 1966; National Science Foundation 1969; Petrowski, Brown, and Duffy 1973).

Table 1 lists alphabetically the top ten universities identified in each of the major studies discussed. Although the domain of graduate education has changed dramatically over the last half century, the same seven institutions appear at the top over the 50-year span of the studies. It would seem that the reputations of Berkeley, Chicago, Harvard, Michigan, Princeton, Wisconsin, and Yale are secure. Nonetheless, the very stability of the prestige of these institutions raises questions. Academe's fondness for ranking institutions and for focusing only on those at the very top betrays a lack of curiosity about educational matters, an indifference to any truth that cannot be reduced to the most or the best.

The Roose-Andersen report indicates that the American Council on Education will not conduct any more studies to assess the prestige of graduate departments. That decision may have been prompted by the methodological, conceptual, and political complexities that surround ranking and reputational studies.

Commentaries and critiques

Considerable attention has been given to the weaknesses of reputational ratings, especially their lack of consensus on the meaning of quality: the definition would seem to vary from rater to rater, from program to program, and from discipline to discipline, making it almost impossible to compare programs and institutions or to develop normative standards. The major implication is that the higher education system is too complex to rank on the basis of one or two dimensions.

Table 1: The Top Ten Institutions Identified in Major Reputational Studies

Institutions are listed in alphabetical order; absolute rank given in parentheses.

Hughes (1925)[a]	Keniston (1959)[a]	Cartter (1966)[a]	Roose-Andersen (1970)[b]
University of California at Berkeley (9)*	University of California at Berkeley (2)*	University of California at Berkeley (2)*	University of California at Berkeley (1)*
University of Chicago (1)*	University of Chicago (6)*	University of California at Los Angeles (10)	University of Chicago (8)*
Columbia (3)	Columbia (3)	University of Chicago (5)*	Harvard (2)*
Cornell (10)	Cornell (9)	Columbia (9)	University of Illinois (10)
Harvard (2)*	Harvard (1)*	Harvard (1)*	University of Michigan (6)*
Johns Hopkins (7)	University of Illinois (10)	University of Michigan (7)*	Massachusetts Institute of Technology (9)
University of Michigan (8)*	University of Michigan (5)*	Princeton (4)*	Princeton (7)*
Princeton (6)*	Princeton (7)*	Stanford (6)	Stanford (3)
University of Wisconsin (4)*	University of Wisconsin (8)*	University of Wisconsin (8)*	University of Wisconsin (5)*
Yale (5)*	Yale (4)*	Yale (3)*	Yale (4)*

Note: An asterisk (*) indicates those institutions common to all four rankings.

[a]Sourse: Magoun 1966.

[b]Source: Morgan, Kearney, and Regens 1976.

Much of the criticism directed at reputational ratings concerns rater bias, which may take several forms. First, overall institutional reputation coupled with insufficient information about particular departments, may produce a "halo effect"; that is, raters who know little about the specific department at an institution may rate it according to their perceptions of the prestige of the institution as a whole. Second, an institution's or a department's reputation may lag behind its current quality and practice. For instance, Cox and Catt (1977) rated psychology departments on the basis of faculty's scholarly contributions to the 13 journals of the American Psychological Association between 1970 and 1975. Comparing their results with the 1970 Roose-Anderson rank-ordered list of psychology departments, they conclude that reputation does indeed lag behind scholarly productivity in the field of psychology and that a reputational survey does not adequately reflect current scholarly accomplishment.

A third form of rater bias is "alumni effect," the tendency of raters to give high marks to their alma maters; complicating the situation, the institutions that produce the largest number of doctorates also produce the largest number of raters. Fourth, an institution's size or age may be reflected in reputational ratings. Finally, even though in some fields nonacademic employers and other "consumer" groups may be more knowledgeable than academicians about program quality, such people are virtually never used as raters in reputational surveys.

Looking beyond methodological criticisms to more substantive issues, opponents of the reputational ranking approach argue that the results of such studies contribute little to a program's self-knowledge or its efforts toward improvement. Moreover, they claim, by focusing attention only at the top level, reputational studies do a disservice to institutions and programs at lower levels, and even to the higher education system as a whole (Blackburn and Lingenfelter 1973; Conference Board of Associated Research Councils 1978; Dolan 1976; Drew 1975; Hartnett, Clark, and Baird 1978; Johnson 1978a; Wong 1977).

While strong objections have been expressed by many (see, for example, Tyler 1972), the strongest objections to reputational studies are expressed by W. Patrick Dolan in *The Ranking Game: The Power of the Academic Elite* (1976). Dolan criticizes the ACE studies because of their inherent tendency to reinforce the status quo and thus to impede innovation and improvement. Neither ACE nor Allan Cartter sought to reform graduate education, let alone higher education, by ranking grad-

uate departments. Indeed, Dolan reports that ACE got involved in "the ranking game" to forestall threatened outside activity in this area: If ACE had not sponsored the 1966 Cartter study, the National Research Council was prepared to conduct its own assessment of graduate programs.

Dolan is particularly skeptical about the criteria used in the ACE studies. The high correlations between quality of faculty and effectiveness of doctoral program, he says, indicate that only one dimension is actually operating to determine the ratings. According to Dolan, subjective ratings of prestige necessarily reflect an elitist and traditionalist view of higher education, a view which discourages or denies diversity, especially as embodied in experimental programs and multidimensional approaches. Thus, large orthodox departments are rewarded for their rigidity and their devotion to scholarship, while the teaching function and undergraduate education are generally ignored. Dolan also believes that, since increasing consumer awareness is the explicit purpose of the Roose-Andersen study, student input should have been incorporated in the ratings.

The careful methodology of Cartter's initial study is implicitly praised by emulators of his approach (see chapters 2 and 3) but strongly condemned by Dolan, who states that the "movement from an interesting opinion poll to the pretense of precise rankings . . . is the most subtle and misleading transition in the studies" (p. 3).

Dolan's final criticisms concern the "uses and abuses" of the ACE reputational ratings, especially in view of the prestige of the sponsoring agency and the systemwide scope of the studies. He fears that they may have an immeasurably adverse impact on individual institutions, administrations, state legislators, and even students, especially as they are used for faculty evaluation and resource allocation. Moreover, Dolan argues that, since the ACE rankings are used by many popular college guides designed to aid prospective students in selecting graduate schools, the studies may even have a deleterious effect on consumer awareness, contrary to their explicit purpose.

By way of contrast, Blackburn and Lingenfelter (1973), in their literature review of reputational studies, defend the ACE ratings on the following grounds:

> 1) Panel bias has been largely eliminated by the careful selection procedures of the ACE studies; 2) subjectivity cannot be escaped in evaluation no matter what technique is used; 3) professional peers are competent to evaluate scholarly work, the central criterion in reputational studies; and 4) although

> not a sufficient condition of general excellence, scholarly ability is necessary for a good doctoral program. (p. 25)

Proponents of quality assessments in higher education may not applaud attempts to rank institutions or departments on the basis of peer ratings, but they do tend to believe that such attempts are inevitable, especially in a period when limited resources dictate the need for careful planning.

Recent quality assessments

In 1976, the Conference Board of Associated Research Councils held a planning session to lay the groundwork for another peer-rating survey of graduate education that would closely parallel the ACE studies (Conference Board of Associated Research Councils 1978). To answer the recurring criticism that raters may be inadequately informed about departments other than their own, the Conference Board proposed to supply information about each program: number of students enrolled, number of doctorates produced in the past three years, and names of faculty members. In addition to peer ratings, data on the career achievements of program faculty and graduates will be collected.

An interesting pilot study of doctoral program quality (Clark, Hartnett, and Baird 1976) was recently conducted under the joint sponsorship of the Council of Graduate Schools (CGS) and the Educational Testing Service (ETS). A sample of 73 departments in three fields—24 in psychology, 24 in chemistry, and 25 in history—were surveyed for the purpose of exploring ways to assess quality. Four major findings emerged: First, dependable and useful information about program characteristics related to educational quality can be obtained at reasonable cost and convenience. Second, between 25 and 30 measures are identified as especially promising. Third, these measures seem to be generally applicable across diverse fields. Finally, two clusters of measures emerge—"research-oriented indicators," including department size, reputation, and physical and financial resources, student academic ability, and faculty publications; and "educational experience indicators," concerned with the educational process and the academic climate (which are rarely considered in quality studies), faculty interpersonal relations, and alumni ratings of dissertation experiences. The variables within each of the clusters are closely correlated with each other but variables from the research-oriented cluster rarely have significant correlations with those from the educational-experiences cluster. Respondents so strongly agreed on the primacy of preparing re-

searchers and scholars that separate analyses based on different program goals were not possible.

The CGS-ETS study also examines peer ratings: their relationship to a broad array of program characteristics, their stability, and the feasibility of using them to rate subdisciplines (Hartnett, Clark, and Baird 1978). The peer-rating component of the study is like that of the ACE study in that "quality of faculty" and "effectiveness of doctoral program" are rated separately but have a correlation of .99; the chief difference is that, in the CGS-ETS study, peer ratings were made by a larger number of faculty members at a smaller number of institutions. The authors report that the resultant rankings are very similar to the 1966 and 1970 ACE lists (with minor variations in the rankings of psychology departments). They do not, however, name institutions.

Subdiscipline ratings, as it turns out, present difficulties: in estimating the extent to which they are subject to departmental halo effect, in dealing logistically with the small number of people involved, and in avoiding the likelihood that raters will have insufficient information on specialties within their own fields. Further, correlations between departmental and subdiscipline ratings are generally so high that collecting subdiscipline ratings in national surveys is probably not worth the trouble. Clearly, however, variations do exist and may be important in individual program evaluations.

The CGS-ETS study used ratings from students and alumni (as well as from faculty) to get supplementary information on departments, and perhaps the most interesting finding is that reputational ratings bear little relation to teaching and educational effectiveness, as revealed by the responses of these groups of raters. Thus, peer ratings seem to be unaffected by the completion rates of graduate students, student perceptions of teaching quality and of the department's concern for students, or the perceived degree of departmental effort toward the career development of junior faculty members. The authors conclude:

> Such data are useful in drawing our attention back to what the ratings are—peers' judgments of the quality of the departments' faculty based largely on scholarly publications. They say little or nothing about the quality of instruction, the degree of civility or humaneness, the degree to which scholarly excitement is nurtured by student-faculty interactions, and so on. In brief, the peer-ratings are not ratings of overall doctoral program quality but, rather, ratings of the faculty employed in these programs, reflecting primarily their research records. No claim has ever been made that the ratings are more than this,

but they have often been interpreted as being more by those who used them. (pp. 1313-4)

As new and multiple indicators of quality are used, and new respondent groups surveyed, the literature on assessment may come to reflect more adequately the scope, diversity, and complexity of the higher education system, even in the graduate domain.

The master's degree

Academic master's-level programs have been ignored in virtually all ranking studies of graduate programs. Providing a rare exception, the Carpenter and Carpenter study (1970) asked the deans of 44 accredited library science schools to rate the overall quality of master's programs, as well as doctoral programs, using Cartter's five-point scale.

When one considers that 311,620 master's degrees were awarded in 1978 and 315,090 are projected for 1980 (National Center for Education Statistics 1980), this lack of information on the quality of the degree seems a decided embarrassment. Several explanations can be offered, however, for the tendency of academic researchers to ignore the master's degree. First, academic departments tend to regard receipt of the master's degree as a step toward the doctorate rather than as a discrete event. As such, the quality of the master's degree is closely linked to the quality of the doctoral program that awards it. Second, some observers (e.g., Dressel and Mayhew 1975; Leys 1956) suggest that the master's degree often serves to screen students for advancement to doctoral candidacy; those students deemed unable to complete the doctorate are awarded the master's degree and gracefully eased out. Finally, in many fields (e.g., education, social work), the master's degree constitutes a license to practice. Any attempt to rank master's programs in these fields would be confounded by the dual academic and professional orientation of such programs (see Glazer 1975).

Under the auspices of the Council of Graduate Schools, attention is being given to the assessment of quality at the master's level through the use of the CGS-ETS multidimensional instrument discussed earlier in this chapter (Clark, Hartnett, and Baird 1976). At the 18th Annual Meeting (Council of Graduate Schools 1978),the responses of 78 CGS member institutions to a survey questionnaire were analyzed to determine the usefulness of individual items on the CGS-ETS instrument in six areas: faculty training and performance, student

ability and experiences, physical and financial resources, judgments about the learning environment, judgments about academic offerings and procedures, and accomplishments of recent graduates. Clark's presentation at the meeting points out that, though evaluations of master's programs have only recently begun, three themes have already emerged from discussions of program review:

> 1. Reviews of graduate programs need to be multidimensional, going well beyond counting number of degrees granted or comparing reputational ratings, if they are to reflect the complexity and variations of graduate education.
>
> 2. Graduate programs should be reviewed in relation to their differing purposes, such as preparing researchers or practicing professionals, meeting local or national manpower needs, or preparing students for doctor's or master's degrees.
>
> 3. Program reviews should lead to the improvement of program quality, rather than focusing entirely on external demands for program accountability. (pp. 213-4)

Emerging interest in the master's degree may be explained by the increasing consumer orientation of evaluations of higher education. After a history of neglect, the master's degree may benefit from what has been learned from quality assessment at the doctoral level.

Summary and conclusions

Reputational studies—with their focus on faculty prestige as perceived by faculty raters, their preoccupation with graduate education and research-related characteristics, and their reliance on similar criteria and methodologies from one survey to the next—have dominated quality assessments of higher education, especially since publication of the first ACE report in 1966. Like disciples following a religious leader, later researchers seem unwilling to question or try to improve on the work of Allan Cartter, even though the same small group of nationally known, long-established, resource-rich universities keep appearing at the apex of the pyramid.

The unfortunate consequences of this situation are perhaps more attributable to the higher education community's competitiveness, the mass media's lust for sensational headlines, and the American public's obsession with knowing who's at the top, than to any fault of the studies themselves. Despite their repeated cautions against aggregating departmental scores to produce in-

stitutional scores and their constant reminders that the ratings represent the subjective judgments of faculty and that they probably reflect prestige rather than quality, scores *do* get aggregated, institutions *do* get compared with one another, and high prestige is translated to mean educational excellence.

As a result, research and scholarly productivity are emphasized to the exclusion of teaching effectiveness, community service, and other possible functions; undergraduate education is denigrated; and the vast number of institutions lower down in the pyramid are treated as mediocrities, whatever their actual strengths and weaknesses.

On the other hand, considering the extent to which the U.S. higher education system has expanded and diversified over the past two decades to accommodate the swelling enrollments caused by both the post-World War II baby boom and the growing demand for postsecondary education, the need to identify and distinguish high-quality programs and institutions is great. The threat of retrenchment in response to shrinking enrollments and tighter resources makes this need even more urgent. Policymakers facing difficult decisions must know what constitutes quality in higher education; in particular, they need to have better information about those programs and institutions that are lower down in the prestige pyramid and thus often fail to be covered in the results of reputational studies.

Assessments of Professional Program Quality

The ACE reputational studies (Cartter 1966; Roose and Andersen 1970) were criticized by some commentators (e.g., Petrowski, Brown, and Duffy 1973) for not including professional programs (though both studies did rate graduate programs in chemical, civil, electrical, and mechanical engineering). The omission was, however, deliberate and can be attributed to two factors. First, professional education is not a specific concern of the American Council on Education. Second, except for engineering and applied sciences, professional education was not a direct beneficiary of the 1958 National Defense and Education Act, which funded the flourishing academic enterprise of the 1960s, thereby providing the impetus for the ACE studies.

Now, however, the climate has changed. Students manifest a "new vocationalism," evidenced in their choice of majors, their aspirations for professional degrees, and their pragmatic attitudes and values (Astin, King, and Richardson 1979). Enrollments in many academic disciplines decline, while applications to professional schools soar. Consequently, the perceived need to rate professional programs and to identify the "top-quality" schools grows more imperative. Margulies and Blau (1973) summarize the situation:

> As professional jobs become scarcer and employers more selective in choosing applicants, the differences among professional schools—in their quality and in their other characteristics—are of growing consequence. Since holders of master's and doctoral degrees have proliferated in the labor market, where he has come from rather than the degree itself, may present an increasingly powerful passport to entry into professions. (p. 21)

As in the graduate domain, the peer-rating, reputational approach to quality has so far dominated assessments of professional education. Further, the methodology of the Cartter and the Roose-Andersen studies is the pervasive model for these assessments. This chapter discusses the two kinds of ranking studies, categorized according to their source: those conducted by the academic community and usually involving assessment of professional schools in several fields, and those conducted by the professions themselves and limited to programs within the single professional field.

Studies by the academic community

An early ranking study of professional education was done by Margulies and Blau (1973) and grew out of a larger study of organizational structures (Blau 1973). Programs in 17 professional fields were ranked on the basis of the number of times that respondents—the deans of professional schools—named an institution (including their own) as among the top five in their field. The Margulies-Blau ratings received widespread attention and criticism, much of the latter being methodological: The overall response rate to the survey was only 36 percent; and in 8 of the 17 fields, institutions were ranked on the basis of responses from fewer than 20 raters.

A year later, the same two researchers received funding from the National Science Foundation and the Russell Sage Foundation to replicate their reputational ranking study—using the same 17 professional fields plus music—with "the aim of maximizing the response rate and thereby increasing the reliability of the rankings" (Blau and Margulies 1974-75, p. 43). In this second study, self-ratings were excluded, and the response rate was increased to 79 percent. The list of leading professional schools, however, remained virtually the same in all fields.

Moreover, Blau and Margulies found that their rankings of the top five institutions agreed with the rankings found in two other reputational studies: the first a ranking study of library science programs in which practicing professionals were used as raters (Carpenter and Carpenter 1970) and the other a (then-unpublished) ranking study of medical schools in which faculty members were used as raters (Cole and Lipton 1977). Blau and Margulies (1974-75) conclude that "the reputations of professional schools in different areas and among different groups of professionals appear to be sufficiently similar to make overall ratings of their reputations meaningful" (p. 46).

The same study looked in more detail at seven fields to see how professional reputation was related to financial resources (total institutional budget and professional school budget) and to "academic climate" (number of books in the library of the professional school and of the institution). As in studies of the correlates of the reputational quality of academic departments (see chapter 1), the size of an institution's library was found to be generally highly correlated with the reputation of its professional schools. Findings with respect to financial resources and size of the professional school library were not so clear-cut: In professional education, reputations "depend on different conditions in different types of professions" (p. 46).

In response to charges leveled against their first study that rankings "engender invidious comparisions and hurt many graduate schools that may not be at the very top of their field" (Blau and Margulies 1974-75, p. 42), the authors contend that since, "after all, professional schools do differ in quality, and these differences concern people becoming affiliated with them, . . . providing information about such differences is a public service" (p. 42). This argument does not really answer the charge: The reputation of a professional school may indeed be harmed by omission from a list of top-rated institutions. Moreover, given the recent proliferation of terminal, professionally oriented programs and the growing diversity of professional education, assessments of quality in this area need to be based more firmly on considerations of possible differences in the goals of different professional programs. Referring specifically to the 1973 Margulies-Blau study, Dolan (1976) comments further on the problem:

> Once again, the assumption that there is a single continuum of quality from the top to the bottom underlies the interpretation. There is no recognition of the fact that quite possibly professional schools with quite different missions and diverse goals could and should be possible, so that quality would mean quite different things in each of the diverse categories. (p 98)

Blau and Margulies (1974-75) express surprise over their failure to find strong significant correlations between the reputations of professional schools and the reputations of institutions in which they are located. (They do not specify how they derived measures to test these relationships, saying only that measures of institutional reputations are based on the Roose-Andersen ratings.) Outside the top five schools, correlations range from only .18 to .35. The obvious question here is: Why should the reputation of a professional school be similar to the reputation of the parent institution? This question becomes especially pertinent when one recalls that "scores" for institutions were derived by aggregating scores for academic departments. Moreover, to anticipate that the reputation of the professional school and of the institution will be closely correlated is to overlook the lesson of the data: That quality should be assessed by field specialization, both in professional programs and in academic disciplines.

Disturbed over the poor showing of California's professional schools in the two studies by Blau and Margulies, the Regents of the University of California commissioned Allan Cartter to conduct a ranking study of professional programs. After Cartter's

death in 1976, the results of the study—which rated programs in law, education, and business—appeared in *Change* magazine (Cartter and Solmon 1977) and in the *UCLA Educator* (Munson and Nelson 1977). Comparing the Blau-Margulies and the Cartter studies, Munson and Nelson conclude that differences between the two sets of rankings can probably be attributed to differences in sample size and selection and in the survey instruments: Blau and Margulies used the deans of professional schools as raters, asking them to name, by recall, the top five programs in their fields; Cartter, consistent with his 1966 ACE study, used deans and faculty members as raters, providing them with a list of institutions to rate on a five-point scale for quality of faculty and on a four-point scale for attractiveness of program (as well as requesting that they indicate how familiar they were with each professional school listed and whether they expected any significant improvement in program). Perhaps the most significant contribution of the comparison is that it helps to clarify the differences between recall-based and recognition-based ratings: Providing raters with a list of schools (recognition) increases the number of contenders for the top and reduces halo effect and alma mater effect (Munson and Nelson 1977), whereas asking raters to name schools (recall) reduces the possibility of prejudicing them by suggesting answers on the survey instrument. These authors further suggest that assessments of professional programs should be done by "a complete sample of experts . . . deans, faculty members, and students at the colleges and universities that supply students to the professional schools being rated, plus the prospective employers" (p. 42).

Another problem connected with assessment of professional education has to do with time lag. In a rating study of programs in educational administration, Gregg and Sims (1972) found "quality of students and graduates" to be the major attribute associated with quality by 725 department chairmen, senior faculty, and junior faculty. Yet, as Blau and Margulies (1974-75) point out, "the fruits of professional training become apparent only years after graduation, so that the quality of a school's program today would have to be judged by the work of its graduates in the 1980s or 1990s" (p. 42).

Though the major product of the Gregg and Sims study (1972) just mentioned was its list of the top 30 of 80 educational administration programs, respondents were asked to indicate what factors they believed determine the quality of an education program; most frequently mentioned was "the provision of relevant educational experiences in the form of intern-

ships and field studies" (p. 82). Reviewing their results, Gregg and Sims assert that "a relatively common value system characterizes scholars in the field of educational administration," concluding that future reputational studies of this type "should utilize appropriate samples of different groups rather than entire populations" (p. 91). Their findings further suggest that the values underlying professional education differ significantly from those underlying graduate education in academic disciplines. By focusing clearly on the link between education and work, the Gregg-Sims study also underscores the importance of referencing quality assessments to the particular field under study, since educational goals and objectives vary substantially from one profession to another.

Studies by the professions

Not all ranking studies of professional education have been conducted by academics; some have been carried out within the professions themselves.

In the field of business, for instance, the staff of *MBA* magazine has conducted two reputational ratings: The first, in 1974, used the recall method ("The 15 Top-Ranked Business Schools in the United States," 1974); the second, in 1975, used recognition ("The Top 15," 1975). Since the list of top institutions was the same in both studies, the attention that the 1975 report gives to slight changes in absolute rankings seems excessive.

Other examples of quality assessment by the professions themselves come from the field of law. In 1976, the staff of *Juris Doctor* magazine surveyed the deans of 167 American law schools and readers of the magazine, asking them to list by recall the top law schools in the country. Responses were received from 58 deans and 1,300 readers. "The results of both polls show clearly that most readers and deans can agree on a group of approximately 20 law schools that today enjoy the highest reputations in the country" ("The Popular Vote: Rankings of the Top Schools," 1976, p. 18). At the same time, the report is full of caveats about possible sources of bias that echo criticisms leveled at academic reputational ratings: Respondents may not be familiar with law schools other than their alma maters and their employers; they may over-rank their own law schools; larger institutions produce more graduates and perhaps more of the magazine's alumni readers; the measurement criteria ("academic quality" and "value in landing good jobs") are vague and subject to various interpretations; numerical ranks

are too absolute, in that a school's being ranked (say) fourth rather than first or second takes on a greater importance than is appropriate; and a program's reputation necessarily lags behind its actual quality.

Quantifiable indicators have also been used to rate law schools. For instance, Lewis (1974) looked at the holdings of law school libraries; although he does not provide a ranking per se, his data could be used to construct one. Similarly, Kelso (1975) included number of volumes in the library as one variable in a "resources index" for ranking law schools; other measures used in the index were the numbers of students and faculty and the ratio of students to faculty, of students to library volumes, and of faculty to library volumes.

Summary and conclusions

As reputational ratings of quality have focused on the domain of professional education, two points have emerged.

First, just as in the graduate domain, although the absolute ranks of institutions vary from one study to another, traditional reputational assessments have consistently identified the same professional schools at the top. The Blau-Margulies (1974-75) and the Cole-Lipton (1977) rankings of medical schools have in common eight institutions among those at the top; three rankings of law schools (Blau and Margulies 1974-75; Cartter and Solmon 1977; "The Popular Vote: Rankings of the Top Schools," 1976) share seven "top" institutions; and three rankings of business schools (Blau and Margulies 1974-75; Cartter and Solmon 1977; "The Top 15," 1975) share six.

Second, reputational studies of professional education require different groups of raters, different criteria, and a different time frame than are usually used in reputational studies of academic disciplines. Further, differences among the professions themselves must be taken into account in designing methodologies for rating professional schools.

The drop in the numbers of the college-age population, the resultant decline in postsecondary enrollments, and the unfavorable academic job market portend that—despite the "new vocationalism" of today's students—professional schools, along with the rest of the higher education community, face difficult retrenchment decisions in the near future. As supply and demand come into closer balance, professional education can benefit from the hard-learned lessons that emerge from studies of quality in the academic disciplines.

Quantifiable Indicators of Quality

In their ceaseless quest for objectivity in assessments of quality in higher education, researchers have explored a variety of quantifiable indicators, used singly and in combination. This chapter looks first at those quantifiable indicators found to be correlates of prestige, as indicated by reputational ratings; then at what is probably the most common quantifiable indicator, the scholarly productivity of the faculty; and finally, at a number of other quantifiable indicators that have been examined in different studies.

Correlates of prestige

Even before the ACE studies, quantitative indexes had been used to rank institutions (Bowker 1964; Eells 1960; Somit and Tanenhaus 1964; Wanderer 1966); and many such studies done later (e.g., Krause and Krause 1970; Packer and Murdoch 1974; Walsh, Feeney, and Resnick 1969) do not involve comparisons with ACE ratings.

Anticipating the results of subsequent research based on quantifiable indicators, Cartter (1966) wrote in the introduction to the first ACE study:

> No single index—be it hize of endowment, number of books in the library, publication record of the faculty, level of faculty salaries, or numbers of Nobel laureates on the faculty, Guggenheim fellows, member of the National Academy of Sciences, National Merit scholars in the undergraduate college or Woodrow Wilson fellows in the graduate school—nor any combination of measures is sufficient to estimate adequately the true worth of an educational institution. . . .
>
> The factors mentioned above are often referred to as "objective" measures of quality. On reflection, however, it is evident that they are for the most part "subjective" measures once removed. Distinguished fellows, Nobel laureates, and National Academy members are selected by peer groups on the basis of subjective assessments, faculty salaries are determined by someone's subjective appraisal, and endowments are the result of philanthropic judgments. Number of volumes in the library, though more readily quantifiable, is a factor of little value in measuring institutional resources unless one can make a qualitative judgment about the adequacy of the holdings. (p. 4)

Despite Cartter's comments about the ultimate subjectivity of all quality measures, the ACE ratings prompted a number of researchers (Adams and Krislov 1978; Clark, Hartnett, and Baird 1976; Clemente and Sturgis 1974; Cox and Catt 1977; Glenn and Villemez 1970; Hurlbert 1976; Johnson 1978a; Knudsen and Vaughan 1969; Lewis 1968; Siebring 1969) to rank institutions on the basis of allegedly objective measures and to compare their results with those of the Cartter (1966) and the Roose-Andersen (1970) surveys.

The list of quantifiable measures of human and material resources that correlate with reputational prestige is enormous. Generally, reputational peer-rating studies reflect research-related variables (Clark, Hartnett, and Baird 1976). Moreover, size alone is a significant correlate (Elton and Rogers 1971; Elton and Rose 1972; Hagstrom 1971), and size is closely connected with research productivity, which also correlates with reputational peer ratings (Drew 1975; Guba and Clark 1978; Knudsen and Vaughan 1969; Wispe 1969). Publication productivity alone is a strong correlate of prestige in some fields (Cartter 1966; Lewis 1968). The prestige of the doctorate institution is closely related to faculty mobility and employment (Crane 1970; Shichor 1970) and to faculty salary (Adams and Krislov 1978; Muffo 1979).

Though the magnitude of all these correlations varies across disciplines, it is generally high enough to suggest that further studies of the relationships between peer ratings of the top domain of the higher education system and the quantifiable indicators mentioned above would be a waste of time.

In a substantial literature review of quality assessment of doctoral programs, Blackburn and Lingenfelter (1973) summarize the findings of studies that attempt to relate reputational ratings to quantifiable indicators. Warning that correlation does not equal causation, the authors list 15 items that are correlated with the 1966 Cartter ratings, as identified primarily by a National Science Foundation study (1969):

1. Magnitude of the doctoral program (number of degrees awarded).
2. Amount of federal funding for academic research and development.
3. Non-federal current fund income for educational and general purposes.
4. Baccalaureate origins of graduate fellowship recipients (NSF fellowships).
5. Baccalaureate origins of doctorates.
6. Freshman admissions selectivity.

7. Selection of institutions by recipents of graduate fellowships (NSF fellowships).
8. Postdoctoral students in science and engineering.
9. Doctoral awards per faculty member.
10. Doctoral awards per graduate student.
11. Ratio of doctorate to baccalaureate degrees.
12. Compensation of full professors.
13. The proportion of full professors on a faculty.
14. Higher graduate student/facutly ratios.
15. Departmental size of seven faculty members or more . . . (this finding is *not* a strict correlate calculated from median scores.) (Blackburn and Lingenfelter 1973, p. 11)

In a study involving a sample of 125 mathematics, physics, chemistry, and biology departments, Hagstrom (1971) found strong significant correlations between departmental prestige (as measured by the Cartter ratings of faculty quality) and a number of quantifiable indicators, including department size (number of faculty members), research productivity, research opportunities, faculty background (including prestige of the doctorate-granting institution), student characteristics (including number of postdoctoral fellows and undergraduate selectivity), and faculty awards and offices. Of special interest is the finding that department size alone accounts for almost one-third of the variance in Cartter's prestige rankings in the disciplines under consideration.

Scholarly productivity as an indicator of quality

Perhaps the most commonly used quantifiable indicator is scholarly productivity of faculty. Despite its popularity, however, the use of this measure is fraught with difficulties. Sixteen years ago, Somit and Tanenhaus (1964) noted that the relatively poor publication records of faculty members at lower-ranking institutions may in part be attributable to their heavier teaching loads, lack of access to adequate library facilities, and other such constraints. It does not necessarily follow, however, that these faculty members are deficient when it comes to training their students in research and scholarship or that the institutions themselves are deficient in teaching and public service.

Even more important, publication productivity may be causally related to prestige and thus unsuitable for use as an independent criterion against which to validate reputational rankings. As Lewis (1968) puts it: "Publication in the leading journals places the name of the institution in the public eye, and it is from continually seeing the name of the institution that

others grant it high prestige" (p. 131). The causal connection may also run in the other direction; that is, the work of faculty members from prestigious institutions may be more readily accepted for publication. Thus, according to Dolan (1976), "the journal publication process itself [is] an exact mirror of the politics of academic prestige trumpeted by the ACE rankings" (p. 42).

A pervasive issue in connection with measures of scholarly productivity is the confusion of quantity of publication with quality, either of the actual product or of its publication source. Frequency tabulations, even those that differentially weight different kinds of publications (with books counting for more than journal articles, etc.), leave much to be desired. As Smith and Fiedler (1971) note, such frequency tabulations make no distinction between worthwhile and inferior books or papers or between prestigious and inferior journals. Yet those studies that attempt to make such a distinction (e.g., Drew and Karpf 1975) have been criticized for excluding new fields and new publications unless they are frequently updated (Brush 1977). Smith and Fiedler also note that scholarly productivity is a more appropriate criterion for some disciplines than for others. Further, they suggest that most attempts to judge the quality of an individual's published work are either superficial, since "rarely is the rater fully acquainted with an individual's writing, it is more unusual for a rater to have read most or all of a scholar's publications" (Smith and Fiedler 1971, p. 226) or logistically impractical, since a thorough reading and content analysis would simple require too much time. Moreover, an individual's productivity changes over time (Bayer and Dutton 1977); men and women differ in their motivations and activity patterns and hence in their scholarly productivity (Astin and Bayer 1972; Bayer 1973). Overall, women are concentrated in teaching-oriented institutions: 68 percent of all women faculty work in lower-tier institutions, compared with 53 percent of men (Shulman 1979). Moreover, women tend to have heavier teaching loads, regardless of the type of institution in which they are employed (Gappa and Uehling 1979). Furthermore, an academic department's productivity changes as funding or age affects the size of its faculty (Drew 1975).

In their review of the literature on the measurement of scholarly work, Smith and Fiedler (1971) emphasize that no criterion measure now available is sufficiently well established to stand alone. They say that the measure that is least contaminated by the prestige factor is *citation count:* the number of

times that a scholar's work is cited in the literature by other scholars. Citations to older research may be given greater weight than citations to more recent research: "According to this rationale, a scholar deserves extra credit if his 15-year old research is still worth quoting" (Smith and Fiedler 1971, p. 228). Nonetheless, the authors note that citation measures still have flaws: At one extreme, significant research may not be recognized for a long time; at the other, the research may be so well known that it is no longer cited by name. Nor it is always possible to distinguish between original and secondary research.[1]

After examining the data from major studies of the correlates of scholarly output, Smith and Fiedler (1971) summarize their findings as follows:

> Quantity of publication is moderately related to individual or departmental eminence, productivity and recognition are moderately related, and citation counts correlate well with recognition and individual eminence. The relationship between citation counts and quantity of publication is less clear . . . [as is] the relationship between citation counts and departmental prestige. . . . The data suggest that citation counts should be compared only within a given field, not between fields. (pp. 232-3)

Other quantifiable indicators of quality

The Blackburn-Lingenfelter literature review (1973) describes and evaluates other quantifiable indicators that have been used to assess quality, including measures of faculty achievement and other traits, student quality, institutional resources, program efficiency, client satisfaction and external viewpoints, and outcomes. Their comprehensive overview is discussed below.

Faculty achievement and other traits—The achievement of faculty, such as degrees and awards, offers another measure of department quality. In addition, some researchers have looked at other faculty traits such as years of teaching experience and travel abroad, but these "are not as useful as measures more closely related to the actual productive work of the faculty, such as scholarly writing or the training of Ph.D.'s" (Blackburn and Lingenfelter 1973, p. 8). Crane (1965) points out that measures of the performance of individual faculty members, as opposed to averaged measures of the performance of all faculty mem-

[1]See Bayer and Folger (1966) and Margolis (1967) for a more detailed discussion of the citation index.

bers within a department, have the advantage of making individual contributions more explicit.

Blackburn and Lingenfelter are less critical of quantifiable indicators based on faculty achievement than are some commentators, failing to note such possible sources of bias as the age of the individual faculty member and the size, age, or eminence of a particular department or institution. Moreover, the awarding of recognition to individuals in academe may be unduly influenced by cronyism and by the existence of old-boy networks.

Student quality—Another approach to evaluating program quality is to examine the quality of the students enrolled in the program. Recalling that Cartter (1966) used the distribution of Woodrow Wilson Fellows among graduate departments as one measure in validating his study, Blackburn and Lingenfelter caution that not all graduate students compete for such awards and that mistakes can occur in the selection process. When Solmon (1976) ranked 50 institutions according to the total number of students with National Institutes of Health (NIH) predoctoral and postdoctoral fellowships enrolled in 1969, he found differences in the enrollment patterns of male and female NIH fellowship awardees. These results suggest that sex, and perhaps racial/ethnic background, should be considered when student achievement is evaluated, as well as when other quantitative measures such as scholarly productivity are used to measure quality, at least until women and minorities become better represented in academe.

Blackburn and Lingenfelter raise the issue of cultural bias in the use of standardized tests to measure the ability of minority-group members, "particularly since equalized opportunity is a desirable societal goal" (1973, p. 9). And, despite the equivocal evidence with respect to role models and student success (Astin 1968; Goldstein 1979), they suggest that "the absence or presence of minority group faculty as mentor-models must be considered in the assessment of a program" (p. 9).

While acknowledging that prestige may contaminate measures of student quality in that good students are attracted to high-prestige programs, Blackburn and Lingenfelter nonetheless assert that "student quality can stand in its own right as a criterion of excellence" because "well-qualified students are an essential element of an excellent program" (p. 8). They also point out that the lack of uniformity in grading standards is a weakness when past academic performance is used to measure student quality, but they ignore other aspects of the problem of student

input: The more capable the student upon entry into a graduate program, the more likely it is that the student's subsequent achievement will be high, whatever the quality of the program. Thus, measures of the achievements of program graduates reveal little about the educational effectiveness of a program unless such input factors as ability, aspirations, motivation, and past accomplishments are taken into account.

Institutional resources—Library holdings are among the most common measures of institutional resources used in quality studies. Like Cartter (1966), Blackburn and Lingenfelter (1973) note the insufficiency of looking just at the number of volumes in the library and the necessity of considering whether library holdings are comprehensive, up-to-date, and easily accessible. Further, they suggest that studies of doctorate education should evaluate the adequacy of laboratories, office space, computer capabilities, seminar rooms, and the like. They caution that, since required facilities differ by discipline, specialists in each field ought to make such evaluations.

Program efficiency—Program efficiency at the graduate level is usually defined in terms of the number of doctorates produced per graduate faculty member or the number of doctoral students enrolled. Moreover, assessments of program efficiency usually involve some kind of cost-benefit analysis. Reviewing the approaches used to assess doctorate quality by measuring program efficiency, Blackburn and Lingenfelter (1973) state that "the ideal index of efficiency in Ph.D. production probably has not been devised" (p. 15). They do, however, offer a list of items that should be included in such an index:

> 1) Enrollment data for students from the time of entry until the termination of their study (with or without a degree); 2) tabulations of individual and departmental activity relative to dissertation committees; 3) tabulation of undergraduate work loads; and 4) tabulation of all instructional activities (seminars, directed readings, etc.) relative to doctoral education. (p. 15)

Client satisfaction and external viewpoints—As Blackburn and Lingenfelter note, "clients" may refer to current students, graduates, or the employers of graduates; all these groups have been surveyed as a means of evaluating quality in higher education. Harvey (1972) reviews the literature on the use of student opinion. In addition, some studies (Bess 1971; Hagstrom 1971) have

looked at faculty morale and satisfaction as indicators of program quality.

Blackburn and Ligenfelter fail to mention the hostility that many academics manifest toward student evaluations, perhaps because they fear that alienated students may be unduly harsh in their judgments. There is also some feeling in the higher education community that surveys of student opinion may amount to little more than popularity contests. Clearly, although student evaluation has been incorporated in some surveys (e.g., Clark, Hartnett, and Baird 1976), the chief "consumers" of higher education have usually not been given an opportunity to make their views known.

Colleges and universities have also been reluctant to give outside observers a voice in quality assessment. Hence, in discussing the intrainstitutional approach to quality assessment, Blackburn and Lingenfelter (1973) maintain that, if external evaluators are used, care must be taken "not to vitiate" the effectiveness of confidential internal self-assessments in enhancing self-evaluation, reducing defensiveness, and providing a "powerful impetus for improvement" (p. 18).

Outcomes—The final group of quantifiable indicators, discussed briefly by Blackburn and Lingenfelter, includes outcomes such as the scholarly productivity of program graduates (a measure that is, of course, subject to the same constraints as measures of faculty productivity and their subsequent employment history), which is usually treated by means of cost-benefit analysis. They mention two major difficulties in conducting cost-benefit analyses in higher education: Lack of information on the actual costs and "exceedingly complex conceptual problems in establishing a valid measure of the social benefits of graduate education" (p. 13). Others have suggested the importance of considering supply and demand in the job market (Shichor 1970) as well as the cumulative effects of time on career patterns, achievement, income, and so forth over the life span (see Mincer 1970). Psacharopoulos (1975) reviews the literature concerning the relationship between institutional quality and subsequent income, asserting that this relationship remains obscure for several reasons. First, in most cases, the size of the samples studied has been "relatively small or too specific for particular groups of people or educational levels" (p. 88). Second, statements about the effect of institutional quality on earnings cannot always be established at a statistically significant level. Third, the question of whether to use independent measures of institutional quality and student ability or whether to "simply use an average measure of student

ability as a proxy for institutional quality" (p. 89) has not been resolved. These reasons would seem to parallel the challenges that continue to confront those concerned with measuring quality in higher education: To ascertain appropriate criteria and to quantify criteria so as to permit comparisons (Blackburn and Lingenfelter 1973).

Summary and conclusions

The literature on the assessment of quality in higher education reveals consensus on a number of needs:

- ☐ Quality assessment must extend beyond the leading 20-to-30 institutions.
- ☐ Multiple indicators should be used.
- ☐ The opinions of consumers (current students, graduates, and the employers of graduates) should be incorporated in program ratings.
- ☐ Some attempt should be made to quantify the social and individual benefits of higher education.
- ☐ More attention should be paid to student learning and growth as the desired outcomes of higher education.
- ☐ Quantifiable indicators must assess adequacy as well as frequency or volume (e.g., library holdings, publications).
- ☐ Different quantifiable indicators are relevant to different disciplines. Moreover, differences between the sexes and among different racial/ethnic groups should be taken into consideration.

Further study is needed to find assessment procedures appropriate to different program purposes and different educational levels. Further research is also needed on transitions from one educational level to another. Urgent questions about which measures of quality are relevant for program improvement and which for policy decisions still have to be answered. Since normative data are necessary to compare programs with one another, it seems desirable to establish ongoing procedures for collecting quantifiable information on physical facilities, faculty quality, and so forth. Most important perhaps, institutions must become more concerned about quality in terms of the development of their students and thus must extend evaluation to consider outcomes other than income, publications, and other easily quantifiable but somewhat superficial considerations. How much weight colleges and universities should assign to the results of external, nationwide research—as compared with their own internal assessments—remains an issue.

Quality Assessment at the Undergraduate Level

That academe has made considerably fewer attempts to assess quality at the undergraduate level than at the graduate level is not surprising: The scope, diversity, and multiplicity of functions that characterize the undergraduate domain make meaningful comparisons among institutions and programs difficult. On the other hand, since undergraduate enrollments far exceed graduate enrollments, valid multidimensional measures of undergraduate quality could benefit both potential students deciding on which college to attend and prospective employers choosing among the graduates of various undergraduate programs.

This chapter first reviews some of the more traditional studies of undergraduate quality, discusses the Gourman ratings (probably the most well-known but at the same time highly questionable ratings of undergraduate education), and then describes an example of the popular college guides offered by commercial publishing houses. The final section discusses the input-environment-outcome model for assessing the quality of undergraduate education, an approach that seems especially promising.

Traditional academic studies

A number of traditional studies rating undergraduate education have demonstrated that colleges differ greatly in their resources (with that term encompassing a multiplicity of factors, human as well as financial). Thus, in a study ranking 119 undergraduate institutions on the basis of multiple weighted quantitative indicators and then comparing each institution's "quality index" score with its library resources, Jordan (1963) found that high-scoring institutions have more library volumes per student and spend more on salaries for library staff than do low-scoring institutions. Moreover, without identifying specific undergraduate schools, Brown (1967) grouped colleges on the basis of eight factors: (1) proportion of faculty with the doctorate; (2) average compensation (salary and fringe benefits) per faculty member; (3) proportion of students continuing to graduate school; (4) proportion of graduate students; (5) number of volumes in library per full-time student; (6) total number of full-time faculty; (7) faculty-student ratio; and (8) total current income per student. These factors are similar to those used to evaluate

graduate and professional programs and thus fail to take into account the special nature of the undergraduate experience.

Astin (1977a) confirms that wealth and resources are unequally distributed among undergraduate institutions, especially in terms of their enrollment of highly able students. He emphasizes that, in light of the differing admissions policies of different types of institutions, equal opportunity and equal access —though "among the most popular cliches in the contemporary jargon of postsecondary education" (p. 8)—may be more myth than reality.

Other traditional studies rate undergraduate institutions on the basis of student achievements. For instance, Krause and Krause (1970) rank colleges according to the number of their baccalaureate graduates who contributed articles to *Scientific American* between 1962 and 1967. Although the authors credit the "potency of small colleges in producing scientists" (p. 134), when one looks at the large schools mentioned in their study, one is forced to conclude that *Scientific American* may be a less scholarly publication than others in which the baccalaureate graduates of larger institutions might publish. Further, graduate study is more appropriately associated with publication than is undergraduate study. Had the results been adjusted to consider graduate school origins rather than baccalaureate origins, the list of larger institutions might well have changed.

Dubé (1974) ranks 100 undergraduate institutions according to the total number of their alumni who entered medical schools in 1973-74. The purposes of the study are not made clear, and the result is a unidimensional, purely statistical portrait; although the absolute ranks might fluctuate from year to year, the same group of institutions would probably emerge at the top in any subsequent rankings. Similarly, Tidball and Kistiakowsi (1976) rank institutions according to the proportions of their baccalaureate graduates who go on to earn doctorates. The same criticism can be applied to both studies: The criterion used is irrelevant for many colleges that do not emphasize preparation for graduate or professional school as the fundamental purpose of undergraduate education. Moreover, the extent of self-selection among "achievers" may be such that their subsequent success can be attributed more to their own abilities and aspirations than to the impact of the college experience (see Anderson 1977; Astin 1963).

In a series of studies, Astin (1965a, 1971; Astin and Henson 1977) has developed a systematic, replicable measure of one aspect of quality in undergraduate education—the selectivity

index, an estimate of the average academic ability of an institution's entering freshmen. The most recent update of the selectivity index (Astin and Henson 1977) uses the SAT and ACT scores of 1973 entering freshmen to estimate the selectivity of all accredited two- and four-year colleges and universities. The authors state that

> educators have a keen interest in selectivity because the folklore of higher education suggests that the more selective institution has higher academic standards than the less selective institution and, by implication, a higher quality education program. Both faculty and administrators are inclined to view the average test scores of their entering freshmen as an index of institutional worth. Regardless of the validity of such views, ample evidence suggests that an institution's selectivity is a good measure of its *perceived* quality. (pp. 1-2)

In short, the more able the student body, the more likely it is that the institution will be perceived as of high quality.

The validity of the latest selectivity index is supported by its correlations with selected institutional characteristics such as tuition and student-faculty ratio. This index has been used not only to rank undergraduate institutions (Astin and Solmon 1979; "Most Selective Institutions of Higher Education," 1978) but also as an independent variable in a longitudinal study of college impact, discussed in greater detail below (Astin 1977b).

Another way of looking at undergraduate quality is exemplified by studies that examine the college preferences of highly able students (Astin 1965a; Astin and Solmon 1979; Nichols 1966) or of students from specific regions and in specific major fields (Astin and Solmon 1979). Astin and Solmon note that "while a popularity measure does not necessarily reflect the average level of academic talent in the student body (i.e., selectivity), it does provide a measure of the institution's drawing power among very bright students" (p. 49). Of course, to assert that the college preferences of highly able students are an indication of the perceived quality of an institution is not the same as asserting that a highly selective institution does a good job of educating its students. Institutional popularity does, however, reveal how much choice exists for—and from—an applicant pool. The authors conjecture that the relative stability over time in the college preferences of highly able students is attributable to the existence of a kind of folklore about higher education quality and that "measures of selectivity and popularity . . . are simply a reflection of the students' ultimate acceptance of this folklore" (Astin and Solmon 1979, p. 50).

Finally, using proportionate numbers of raters from 20 per-

cent of all U.S. colleges and universities, *Change* magazine (Johnson 1978a) conducted a reputational study, based on rater recall, to investigate three senses of the term "leadership." The study found that raters from all types of institutions agree on a list of those institutions "leading" in national influence, and, in fact, confirm "the traditional cluster" of top-rated institutions identified in previous graduate and professional rankings. They found, however, that when asked to list those institutions that were leaders in the sense of being innovators, "respondents from all types of four-year institutions cited liberal arts colleges more than other types of institutions"—especially those liberal arts colleges with highly selective admissions policies and those that are the leading producers of baccalaureate graduates who go on to the get the doctorate—whereas two-year college respondents "mostly cite community colleges" (p. 51).

The Johnson study also examined the results with respect to geographic proximity. According to this analysis, even though institutions of all types are unlikely to cite institutions within their own state when asked to name national leaders or innovators, they are likely to mention such institutions as having a major influence on their own programs, especially those institutions belonging to the same Carnegie category. Further, community colleges are most likely to manifest this regionalism. The *Change* study calls into serious question, then, the utility of the simple rank orderings reported in past reputation surveys. It suggests that broader issues of education involving undergraduate as well as graduate programs ought to be carefully considered whenever institutions are ranked. Johnson (1978a) states, "the structure of American higher education is far too complex to be understood in relation to any single academic procession" (p. 51).

Indeed, another means of rating, and essentially ranking, almost the entire higher education system is the Carnegie classification (Carnegie Council on Policy Studies in Higher Education, 1976) used by Johnson (1978a) and others. Despite the alleged objectivity with which institutions are listed in six broad categories, it is clear that the most prestigious are included in the subcategory "Research Universities I" for graduate study and "Liberal Arts I" for undergraduate study.

In summary, traditional academic studies, whether reputational or based on quantifiable indicators, tend to stand as discrete entities. Systematic investigation of how to measure quality and of what quality means in the heterogeneous undergraduate domain has so far been lacking.

The Gourman ratings

Though probably the best known of undergraduate ratings, the Gourman ratings (1967, 1977b) are idiosyncratic and unreplicable. Neither report fully explains the methodology used to derive the ratings. What *is* revealed, however, may help to account for some of the odd results.

In the 1967 ratings, 1,187 four-year colleges were scored on two sets of variables: strength of the institution's academic departments and quality of nondepartmental areas. The scores were expressed as letter grades corresponding to the College Board scale: A=800, B=600, C=400, D=200. Then, variable scores in each set were averaged to produce a numerical "average academic departmental rating," and "average nondepartmental ratings," and an overall "Gourman rating" for each institution.

Although the Gourman index has been used as a basis for other studies (e.g., Solmon 1975), many of Gourman's assertions are highly questionable. Thus, he rates "older" college faculties more highly than "younger" ones on the grounds that "a minimum of ten years after college graduation is necessary to produce an excellent teacher in the classroom" (p. xiii) but offers no evidence to substantiate this claim. Moreover, equal weight is given to ratings of a college's alumni association, faculty effectiveness, public relations, library, and athletic-academic balance, even though common sense suggests that these factors differ considerably in the magnitude of their contributions to institutional quality. Finally, Gourman reveals a bias toward large institutions, tending to rate large public institutions more highly than smaller liberal arts colleges (Webster 1979).

The 1977 Gourman ratings use a format identical to that of the 1970 Roose-Andersen study of graduate programs to rank only 68 undergraduate programs, as well pre-medical and pre-law programs in the U.S., and foreign/international universities and professional schools. Again, no information is given as to how ranks and scores were derived, what factors were considered, or how these factors were weighted. Supposedly, these methodological matters are dealt with in "supplemental reports" on institutions; however, no such reports have ever appeared.

An example of a popular college guide

In addition to academic studies of undergraduate quality, a number of guides to undergraduate colleges are available from com-

mercial publishing houses, *Hawes Comprehensive Guide to Colleges* (1978) being a good example. "Based on research data—not opinion" (p. xi), this publication rates almost every two- and four-year college in the country.

Perhaps the most interesting aspect of *Hawes Guide* lies in its implicit view of the purposes of undergraduate study and the missions of undergraduate institutions, a view clearly at odds with that of most educators and which raises a compelling question: Does *Hawes Guide*—and others of its ilk—reflect what prospective students and their parents, as well as others outside the academic community, want from higher education? Consider, for example, the following criteria used in *Hawes Guide* as measures of undergraduate quality[1]:

1. "Social prestige" ratings, based on the number of an institution's graduates listed in the current edition of the *Social Register*. This information is given so that the prospective student may know "the extent to which the sons and daughters of America's upper class—its richest, oldest, most socially prominent families—go to that college" (p. xi).

2. "Social achievement" ratings, based on the number of an institution's graduates listed in the current edition of *Who's Who in America*. Supposedly this information indicates "how likely this college is to help a student achieve high status later in life largely through his or her own abilities and efforts" (p. xii).

3. Consumer ratings: Some institutions are labeled "best buy," "better buy," and "good buy."

4. "Faculty salaries" ratings, said to be "one very basic indicator of the college's academic quality" in that "a college with higher faculty salaries will in general attract more highly qualified professors" (p. xii). (Faculty salaries, however, are not adjusted for geographic, or other cost-of-living differences.)

5. "Expense" ratings that indicate "the level of dorm-student expenses" (p. xii).

6. "Admissions" ratings: "hard," "selective," and "easy."

The fallacy inherent in using mention in the *Social Register* is readily apparent; many people are listed in this publication by virtue of their parents' or their spouse's status. Moreover, such a criterion seems inappropriate for higher education in a demo-

[1]Items 3, 5, and 6 in this list are not used in calculating the rank of institutions.

cratic society, particularly in a period when concern over equal access and affirmative action runs high. Similarly, mention in *Who's Who* is a questionable criterion, even though some researchers (e.g., Tidball 1973) also have asserted that such mention is related to the quality of one's undergraduate institution. The primary difficulty is that this criterion confounds the impact of the undergraduate institution with the abilities and efforts of the individual, who might very well achieve such mention whatever his or her undergraduate origins. Nonetheless, it is interesting to note that those institutions that rank at the top in "social prestige" and "social achievement" tend to be the same institutions that rank at the top in reputational studies of graduate education. The most reasonable explanation for this correspondence is that those institutions that are most highly visible and prestigious are the same institutions that tend to attract affluent and highly able students.

The input-environment-outcome model

As has been suggested, the principal drawback to assessing an undergraduate institution's quality on the basis of such factors as selectivity or alumni achievements is that these factors tell us nothing about the contribution of the institution itself. That a highly selective institution tends to produce high-achieving graduates is not necessarily to the credit of the institution or its programs; these individuals might well have gone on to be high achievers whatever their undergraduate origins. Similarly, such institutional resources as highly credentialed and highly productive faculty, a comprehensive and up-to-date library whose materials are easily accessible to students, and superior laboratory, computer, and classroom facilities should be regarded as indicators of quality only insofar as they can be proved to have desirable effects on the development of undergraduates. Most educators would surely agree that the chief purpose of undergraduate education is to bring about or to facilitate some kind of positive growth in students. Thus, assessing the degree to which different institutions contribute to such growth provides a sound basis for comparing the quality of different undergraduate institutions. For this reason, the input-environment-outcome model represents the most promising approach to such quality assessment.

In this model (see Astin and Panos 1969), *input* is defined as what students bring with them to college: their prior knowledge, abilities, aspirations, and motivation, as well as such background characteristics as sex, race/ethnicity, and socioeconomic

status. The *environment* comprises not only an institution's educational programs but also its other resources (including extracurricular activities) and characteristics to which students are exposed. The *outcome* component of the model can be described according to three dimensions, all of them involving *changes* in students (Astin 1974; 1977b):

1. Type of outcome: cognitive or intellective changes (e.g., in reasoning ability) versus noncognitive or affective changes (e.g., in values and attitudes).

2. Type of data (that is, the type of information used to assess cognitive and noncognitive outcomes): psychological data, which relate to "the internal states or 'traits' of the individual" (Astin 1977b) versus behavioral/sociological data, which relate to observable behavior.

3. Time dimensions: long-term versus short-term effects. The relevance of any measure of input or of environment depends upon what outcomes are being evaluated (Astin 1974).

Moreover, the findings from a study of college students several years after their graduation (Solmon and Ochsner 1978) underscore the importance of distinguishing between the short-term and the long-term effects of college. Whereas Astin (1977b) found that student values tend to decline over the college years (in that smaller proportions of seniors than of freshmen rate as essential or very important a number of life goals), and that these declines are most marked with respect to status needs and business interests, Solmon and Ochsner reported that, several years after graduation, interest in certain life goals (e.g., being very well-off financially) had once again increased, suggesting that "the effects of college on values and life goals do not endure long after graduation" (1978, p. 2).

The particular utility of the input-environment-outcome model is that it permits the researcher to apply statistical control for student input variables and thus to assess the actual contributions of environmental variables (i.e., the college experience) to the outcomes under consideration. Thus, the impact of different colleges and different college characteristics on student development can be isolated.

There are several reasons why this model has not been used more widely in studies of academic quality. The first is a lack of consensus within the academic community on the proper goals and objectives (i.e., desired outcomes) of higher education. Second, even when goals and objectives are agreed upon, they are often stated in vague or abstract terms (e.g., "to make the stu-

dent a well-rounded person," "to improve critical-thinking ability") that are difficult to operationalize. Third, the model requires a more sophisticated and elaborate methodology than is involved in (for instance) counting faculty publications or number of volumes in the library. Most important, the design should be longitudinal; that is, students must be surveyed at the time of college entry to determine their input characteristics and then followed up at some point after exposure to the college environment (e.g., two years, four years, ten years after college entry) to assess change.

Nonetheless, the model has occasionally been promoted as a means to assess the graduate domain (Blackburn and Lingenfelter 1973; Clark, Hartnett, and Baird 1976; Conference Board of Associated Research Councils 1978). More frequently, it has been applied in studies of undergraduate education (Astin 1965b, 1970, 1974, 1977b; Astin and Panos 1969, 1971) as a means of assessing "value added" by college attendance.

Those who argue for the superiority of such a model do not always agree on the assignment of variables among the three components. For example, Blackburn and Lingenfelter (1973) regard "faculty characteristics" as an input variable, whereas Astin (1977b) and the Conference Board of Associated Research Councils (1978) consider this variable to belong to the environment component. Perhaps these differences are attributable in part to inherent differences between graduate and undergraduate education: The goals and objectives of graduate education tend to be relatively clear-cut and more widely agreed to, whereas the goals of undergraduate education are more numerous and more diverse and thus require that careful theoretical rationales be constructed prior to evaluation. Indeed, to circumvent the problems involved in specifying goal accomplishments in higher education, Cameron (1978) proposes a model for measuring the concept of "organizational effectiveness" using nine criteria; such an approach investigates the environment of the system rather than its outcomes.

Kerr (1978) reminds us that, in college, "what happens along the way is often more important than the purpose of the journey" (p. 167). The impact that an undergraduate institution has on the development of its students should surely be regarded as a fundamental measure of its quality. As more research is conducted on college impact using the input-environment-outcome model and focusing on student growth or "value added" as a major consensual goal (or, of necessity, specifying other goals and objectives so that the extent to which they are achieved may

be assessed), the meaning of quality in undergraduate education will assume more appropriate scope and diversity than is possible from traditional approaches borrowed from studies of the graduate domain.

Summary and conclusions

The academic community has conducted relatively few comparative assessments of undergraduate programs. Moreover, because different criteria are used from one study to the next, the assessments that have been done have produced rankings that are not comparable. Unlike reputational rating studies in the graduate and professional domains, ranking studies at the undergraduate level do not produce identical lists. Perhaps because of its diversity, the undergraduate level inevitably assumes varied hierarchies according to the criteria used to rate it. In the absence of consensus on the goals and objectives of undergraduate education, studies that focus on student change or "value added" and that apply the input-environment-outcome model (which also provides a useful framework for assigning quantifiable indicators to different components) may be most valuable.

If comparisons among undergraduate institutions must be made, different types of institutions (e.g., two-year and four-year colleges) should probably be considered separately, and cognizance should be taken of the uneven distribution of higher education institutions within and among states. Moreover, although the point is rarely discussed, institutional assessments may not adequately reflect the existence of especially strong—or weak—departments. On the other hand, "departmental quality" may represent too narrow a criterion for assessing undergraduate education, where students are exposed to a broader range of disciplines than is true of graduate students and where other environmental characteristics may play a critical role in enhancing or detracting from the undergraduate experience.

Undergraduate education presents a challenge beyond that of quantifying or standardizing criteria so as to permit comparisons among programs and institutions: That challenge is to find criteria appropriate to the size, heterogeneity, and multiplicity of functions that the undergraduate experience encompasses.

Other Dimensions and Concerns in Quality Assessment

Assessments of quality in American higher education can be described as having either an internal or an external focus. Examples of the former include the ACE ratings, while the latter type of quality assessment comprises the accreditation process and the state program review process.

Thus far, the discussion has centered on the issue of quality in higher education as assessed, described, and critiqued in the research (i.e., academic) literature. These internal assessments —e.g., the ACE surveys, the Blau ratings, single-discipline reviews, and the correlational studies generated by all the preceding—constitute a literature intended primarily for an academic audience. To be sure, these studies are public documents, and some are reported in the mass media. Parents and prospective students, both undergraduate and graduate, may look through these materials in their efforts to find the "best" information on which to base enrollment decisions. Nonetheless, these documents are of primary interest to academics and relate most to the "private life" of higher education (see Trow 1975). The tendency to view these ratings as absolute or ultimate assessments of program and institutional quality, against the warnings of both the researchers and the critics, is likely to increase when such reports are used by the general public.

What, then, of external types of quality assessment? What is the nature of such activity? How does it differ from internal assessments? While the interest in, and the furor created by, the ACE and similar ratings during the heyday of postwar academic expansion seems to have subsided as higher education enters the "no-growth" era of the 1980s and 1990s, interest in accreditation (the oldest form of quality assessment) and state program review (the newest form) is growing, spurred by two major trends: increasing governmental concern about the financial accountability of higher education; and increasing public concern about the outcomes or benefits of college attendance. Displacements in the job market for college graduates, societal commitment to the goal of equal educational opportunity, institutional dependence on direct and indirect federal support (research grants and student aid payments), and emphasis on consumer protection have all contributed to the current interest in external assessments of quality in higher education.

Accreditation

What is it about accreditation that assumes (or assures) institutional quality and inspires the faith of college-bound students, their families, and government agencies? Even though accreditation standards are not widely understood by the general public, students and their parents look to accreditation as an indicator of institutional quality and stability, and institutions respond to these concerns by listing their affiliations with various accrediting bodies in their promotional literature. Accreditation is, in most instances, a prerequisite for participation in federal aid programs, both for institutions and for students (that is, students must be enrolled in accredited institutions to receive federal financial aid). Yet how strong is the relationship between accreditation and quality? And what are the attributes of institutional quality as defined in the literature on accreditation?

Accreditation and quality—The relation between quality and accreditation is made explicit in the statements of definition and purpose offered by experts in, and representatives of, the field. Some examples of their views follow:

- Kenneth Young (1976a), president of the Council on Postsecondary Accreditation (COPA), the national nongovernmental coordinating organization for accrediting agencies, says that "if accreditation can be defined in 25 words or less that definition would be: 'Accreditation is a process that attempts to evaluate and encourage institutional quality'" (p. 133).

- According to Harcleroad and Dickey (1975), accrediting serves as "the major factor in quality control for our institutions of higher education and for various professional and specialized programs" (p. 7).

- Patricia Thrash (1979), of the North Central Association, states that accreditation "provides an assurance of . . . educational quality and integrity . . . to the educational community, the general public, and other agencies and organizations" (p. 116).

- The Advisory Committee on Accreditation and Institutional Eligibility (U.S. Department of Health, Education, and Welfare 1977), a federal advisory panel, asserts that the federal government uses accreditation as an eligibility criterion for participation in federal programs because accreditation provides "a reliable authority concerning the quality

of training offered by institutions and programs." (p. iii, see also Trivett 1976, pp. 8-19).

Promoted as an attribute of institutional quality, accreditation—because it is essentially a binary process—may actually impede true assessments of institutional quality. Accreditation provides for an assessment of institutional performance against institutional objectives or against other (baseline) standards; and, operationally, an institution or program either is, or is not, accredited. In contrast, quality (like wealth, beauty, and wisdom) exists on a continuum.

While the accrediting community has been active in asserting the relation between quality and accreditation, it has been less precise in defining the actual attributes that make for institutional and program quality, probably because of the cherished diversity of the American higher education system (which does not lend itself to uniform operational definitions) as well as the consensual nature of the attributes of quality: We all (think we) know what quality is when we see it, but we have difficulty describing it for others.

Accreditation's historical movement from quantitative to qualitative evaluation suggests that the accreditation process is primarily a criterion-referenced assessment.[1] The regional associations' self-study guides and accreditation documents describe the accrediting process as the assessment of an institution in terms of its stated purposes and objectives. Yet some accrediting agencies currently do provide quantitative guidelines, and many are indeed interested in quantifiable data that help to describe institutional attributes and resources (Petersen 1978). The ambiguity of some of the criteria would appear to give accrediting agencies flexibility with respect to enforcing standards; the diversity of the American system of higher education would appear to require it.

Accrediting criteria—The regional and professional associations, whose basic task is to insure that minimal standards are operationalized, have articulated certain principles and criteria, often referred to as standards or guidelines, which are promoted to be attributes of institutional and program excellence or quality.

Reviewing the published standards and guidelines of both regional and professional associations, Petersen (1978) con-

[1]Readers interested in the history of accreditation are referred to Dickey and Miller (1972), Selden (1960), and Harcleroad (forthcoming).

cludes that "there is such a wide variety [of standards] among agencies that almost any blanket conclusion or generalization is suspect" (p. 306). Harris (1978) offers a somewhat different opinion. In a report prepared for COPA, he identifies seven criteria as being critical characteristics of an "accreditable" institution.[2]

1. Goals and objectives: Because institutions are evaluated on the basis of their own purposes rather than by external standards, they must have explicit, comprehensive, and consistent goals and objectives that are subject to periodic review and revision.

2. Governance, leadership, and structure: A basic premise of accreditation is that faculty possessing proper credentials will be significantly involved in designing curricula, setting graduation requirements, and evaluating students; faculty, therefore, will maintain academic standards because an appropriate structure of academic and administrative checks and balances exists to monitor effectively the institution with respect to its purposes, programs, curricular planning, and degree requirements.

3. Validity of degrees: Student achievement is commensurate with the general meaning of degrees awarded, and the institution has a systematic means to assure that students meet the letter and the spirit of degree requirements.

4. Adequate resources: Adequate human, physical, and fiscal resources, as judged by academic peers, exist to accomplish stated goals and objectives.

5. Stability: The prevailing values of the academy are best represented by institutions that display evidence of stability and permanence.

6. Students and programs: Student needs, interests, and aspirations are reflected in institutional programs, and those services logically related both to the institutional mission and to student needs are provided.

7. Integrity: Institutional integrity is reflected in explicit goals and objectives; full disclosure of codes, rules, and practices; sound fiscal management; ethical recruitment and promotion practices; consistent application of institutional codes;

---

[2]Harris (1979) focuses on "accreditable" instead of "good" because the former term is the "more operational adjective," and because of the membership component in the accreditation process: i.e., "accreditation means that an institution makes itself amenable to the criteria and the procedures of the association in which it seeks membership" (p. 63).

and continued monitoring and self-assessment of institutional behavior and practices against stated goals and objectives.

Harris (1978) suggests that accreditation policies reflect "the conventional wisdom of the academy [at any point in history] about quality" (p. 62). Yet current developments—such as nontraditional education, the increasing significance of accreditation in the quest for federal dollars, and the shift, at all degree levels, from a seller's to a buyer's market—pose a number of challenges to the "conventional wisdom" regarding quality and accreditation.

Troutt's (1979) textual analysis of the published criteria of the six regional accrediting associations reveals five criteria that "claim some association with quality assurance. . . . Most regional associations suggest a relationship between institutional quality and criteria for: (1) institutional purposes and objectives; (2) educational programs; (3) financial resources; (4) faculty; and (5) library/learning resources" (p. 200). Troutt identifies three basic assumptions underlying the criteria that the regional associations promote as being related to institutional quality. First, judgments about quality should be based on inferences from specific conditions rather than on a direct evaluation of student performance. Second, no common benchmarks exist for measuring institutional quality. Finally, accreditation criteria equate higher education with a production process. These three assumptions contrast sharply with those of educational researchers (e.g., Astin 1977b; Dressel 1978) who assert that quality judgments should be based on an assessment of student outcomes, that common benchmarks do exist, and that the production model is neither the only, nor the best, model for describing higher education (see Clark et al. 1972; and Walsh 1973).

Graduate program accreditation, in contrast to general institutional accreditation as coordinated by the regional associations, is somewhat more specific about the attributes of program quality. Graduate education is seemingly a more sacred bastion than undergraduate education. Anderson (1978) observes that while the "higher education establishment could tolerate wide diversity and lesser quality in undergraduate programs and even at the master's level . . . it registers deep concern when the quality of the doctorate is diluted" (p. 279). Andrews (1978) asserts that there is an inverse relationship between enrollments by degree level and concern for program quality in higher education: Graduate and professional programs, which enroll the smallest number of students, have historically been the focus

of the debates on quality, while lower-division, undergraduate, and vocational education have generally received little attention in such discussions. Our survey of the literature confirms this contention: Articles and documents on graduate education and graduate rankings outnumber those on undergraduate programs by a ratio of roughly six to one.

State program review

The state role in higher education has changed considerably during the last 15 years: from passive purveyor to concerned underwriter. Similarly, the role of the program review process has changed in response to a number of recent developments: increased financial and political pressures for the efficient use of resources, the proliferation of degree programs at all levels, the shrinking job market for degree-holders. Although other purposes are attributed to the review process (e.g., to eliminate unnecessary program duplication, to assure quality), the term "accountability" not only best describes its rationale but also subsumes the other purposes attributed to it (Barak and Berdahl 1978).

The state perspective on quality—To understand the place of program quality assessment in the state review process, one must first be aware of the historical state perspective on quality.

Like the higher education community in general, the state would seem to have rather traditional notions about institutional and program quality (see Halstead 1974, chapter 6). During the postwar period of rapid growth in higher education, the states viewed quality as manifesting itself primarily in criteria established by the academy: i.e., students with high test scores and faculty with doctorates, research grants, and publications. These are the attributes of quality that receive most attention in the literature of the period (e.g., Berelson 1960; Committee of Fifteen 1955). The states purchased (or created) higher education facilities for the benefit of their citizens, and the states "bought" the value system of the academic community. The emphasis on student and faculty credentials as attributes of institutional quality was a response to market factors during the late 1950s and most of the 1960s, when "high-quality" students and faculty were in short supply.

Halstead (1974) indicates that state planning agencies have generally accepted responsibility for providing the leadership to improve the quality of public higher education. Until very recently, however, the states viewed program quality as depending

almost entirely on the input characteristics of students and faculty. The traditional state perspective is exemplified in the following statement from the 1960 *California Master Plan for Higher Education:*

> The quality of an institution and that of a system of higher education are determined to a considerable extent by the abilities of those it admits and retains as students. This applies at all levels—lower division, upper division, and graduate. It is also true for all segments, but the emphases are different. The junior colleges are required by law to accept all high school graduates (and even some nongraduates under some circumstances); therefore the junior colleges must protect their quality by applying retention standards rigid enough to guarantee that taxpayers' money is not wasted on individuals who lack the capacity or the will to succeed in their studies. If the state college and the university have real differences of function between them, they should be exacting (in contrast to public higher education in most other states) because the junior colleges relieve them of the burden of doing remedial work. Both have the heavy obligation to the state to restrict the privilege of entering and remaining to those who are well above average in the college-age group (California State Department of Education 1960, p. 66).

As this statement makes clear, the California master plan is based on a meritocratic model of program and institutional excellence in that it provides greater resources and opportunities for the academically endowed while regarding those students "who lack the capacity or the will to succeed" as antithetical to institutional quality.

More recently, the states have moved beyond this perspective, expanding their focus to include educational process (e.g., the provision of educational services, the impact of educational experiences) as well as student and faculty input characteristics as manifestations of quality. This shift in perspective is in large part a response to the demands from a number of constituencies for an accounting of (1) the resources allocated to public postsecondary education and (2) the availability and distribution of educational opportunities and benefits to various clienteles (see Callan 1978).

Kerr (1973) was among the first to describe the need/access versus quality/excellence debate (which centers on the availability and distribution of educational opportunities and benefits) from the standpoint of the states, warning that they would find it difficult to satisfy the academic community's heightened expectations for program expansion and quality improvement

and at the same time accommodate the increasing number of high school graduates (and returning adults) with degree aspirations. Indeed, the state interest in, and responsibility for, educational access and opportunity at all degree levels may well conflict with traditional notions of academic quality.

State program review criteria—From 1960 through 1975, not only the number of state agencies but also their capacity to conduct academic program reviews increased significantly. Using data from the U.S. Office of Education (Martorana and Hollis 1969) and the Education Commission of the States (1975b), Barak and Berdahl (1978) document a 105-percent increase (from 19 to 39) in the number of states with higher education coordinating or governing boards that have program review authority over both new and existing programs. The number of states with governing or coordinating agencies that have authority only to approve new program proposals also increased over this period, from four to eight. Barak and Berdahl note, however, that an agency with legal review authority may not, for a number of reasons, exercise that authority, whereas in some states a review authority that does not exist in law may be exercised by other means and by other agencies (e.g., legislative budget reviews).

Developing appropriate criteria for program reviews has not been easy: The process is as political and volatile as any activity inside or outside the academy (Barak and Berdahl 1978; Hill 1978; Hill et al. 1979; Mingle 1978). The Task Force on Graduate Education of the Education Commission of the States (1975a) recommends ten factors to be considered in the program review process. These factors are listed below; the figures in parentheses indicate the number of states—among the 27 that currently conduct some sort of program review or that have established procedures for review—that use the criterion:

1. Number of program graduates in each of the five preceding years (15)
2. Student enrollment (matriculation and retention) (12)
3. Size of classes and cost of core courses (5)
4. Cost per program graduate (i.e., per degree awarded) (9)
5. Faculty workload (2)
6. Program quality, as reflected in (a) reputation, (b) faculty qualifications, and (c) the employment experience of program graduates (3)
7. Comparative analysis of the production of program

graduates from similar types of programs in the state, the region, and the nation (3)

8. Economies or improvements in quality to be achieved through program consolidation or elimination (3)
9. General student interest and demand trends (10)
10. Appropriateness of the program, given the institutional mission (10)

The most frequently used criteria are measures of productivity, costs, and the compatability between program and institutional mission (Barak and Berdahl 1978).

Barak and Berdahl (1978) identify nine states that consider program quality in the review process. It is interesting, but not surprising, that the states generally have not developed new procedures for assessing program quality; rather, they tend to adopt the procedures and criteria articulated by accrediting agencies and educational researchers: e.g., student characteristics, faculty qualifications and research productivity, and peer review. Review procedure guidelines, agency policy statements, and evaluation committee reports indicate the extent to which traditional measures of quality have been accepted in the state program review process. For example, a 1973 policy statement of the New York Regents proclaims that

> the attributes of [the] quality of a program are widely known and accepted. Among these are the level of faculty research and scholarship; the effectiveness of and attention to teaching and counseling by the faculty; the caliber of students; the caliber of dissertations; the adequacy of laboratory, library, and other related facilities; the presence of supporting and related programs. (Regents of the University of the State of New York 1973, pp. 17-8)

Following the lead of the Regents, the report of the New York Chemistry Program Evaluation Committee states that the factors "of central importance to the committee" as measures of program quality were "the quality of the faculty, the research interests of the faculty, and the quality of the students (State of New York, Chemistry Program Evaluation Committee, n.d., p. 5).

While a number of states consider productivity factors in the review process, only the Florida Regents see productivity as being directly related to program quality:

> It would be impossible to conduct a thorough investigation of every program every year. The use of degree productivity as a means of identifying programs to be evaluated rests on the assumption that with the exception of professional programs

> such as medicine and law, degree productivity is the best single index which correlates meaningfully with the enrollments of majors in the program, student demand, [the] job market for graduates, and [the] quality of the program. (Florida Board of Regents, March 8, 1974; cited by Barak and Berdahl 1978, pp. 62-3)

This view is indeed unusual.[3] Although productivity issues are a concern in the review process, few states have broken new ground by expanding the conceptualization of quality criteria in the manner articulated by the Florida Regents. In summary, most states continue to view quality from the traditional perspective, focusing on process variables (e.g., faculty, institutional resources) and input variables (e.g., student characteristics).

Summary and conclusions

Historically, the states have financed higher education and left the issue of quality assessment and management to the academic institutions, as implemented by the accreditation process. In the past 15 years, public concern for excellence/quality in higher education has been affected by a number of factors: (1) reduced demand for higher education, and the financial consequences, including accountability, which have accompanied it; (2) federal incentives, such as the 1202 legislation, which promotes statewide and regional planning and coordination; (3) the postwar transition of higher education from option/opportunity to entitlement, formalized by the Basic Grants (BEOG) legislation of the Educational Amendments of 1972; (4) increasing concern for consumer protection; and (5) growing emphasis on the outcomes and benefits of college attendance, stimulated by the equal opportunity concerns and the job market displacements of the 1970s. These factors, and others, have served to focus renewed attention on accreditation and new attention on state program reviews as "public" or external assessments of quality in American higher education.

Accreditation has two characteristics that distinguish it from other forms of quality assessments. First, accreditation focuses on an institution's capacity to achieve, and the extent to which an institution does achieve, articulated goals and objectives. Second, accreditation assessments are not competitive: i.e., in-

[3]The use of productivity measures as a yardstick for assessing program quality and as a vehicle for identifying those program requiring more comprehensive review proved to be controversial and has since been modified.

stitutions are not compared and ranked. State program reviews, primarily concerned with resource allocation, are unique in that they address issues pertaining to finances, access and opportunity, service to client populations and to the commonweal, and productivity. Taken together, accreditation and program reviews add other dimensions and other concerns—public concerns—to the discussion of quality in higher education.

Conclusions

The previous chapters have reviewed the literature to illuminate the question: What is quality in American higher education? It would appear that the definition of quality varies with the context, depending on who is doing the assessment, by what means, and for what purpose.

In academic studies, usually conducted by researchers from the higher education community, assessments have focused on identifying the "best" institutions (or graduate departments). Whether based on peer review or on the application of a set of traditionally-used quantifiable indicators (which generally correlate highly with each other and with peer ratings), such assessments simply ignore about 99 percent of the institutions that constitute the nation's higher education enterprise. Moreover, as critics have pointed out, these rankings serve to reinforce the hierarchical structure of the system, in that those few institutions, departments, or professional schools at the top of the pyramid continue to capture what may be more than their fair share of scarce resources (including highly able students). Thus, their prestige is further enhanced, while the incentive to improve their educational programs may be reduced.

From this review of the quality literature, certain conclusions emerge as to how quality in higher education might be better defined and how methods of assessing quality might be improved.

First, quality assessments must be referenced to departmental or institutional goals and objectives. Although academics may find it hard to agree on the proper goals and objectives of higher education, some specification of desired outcomes is required as a first step in the assessment process.

Second, the diversity of American higher education must be recognized and accepted rather than (as is too often the case) simply paid lip service. Different institutions and programs serve different constituencies and have different goals and objectives. To measure them all by the same yardstick is to do a disservice not only to the higher education system but also to prospective students and to the public as a whole. Rather, those concerned with assessing quality must be more flexible, more willing to try a variety of quality measures or criteria that may be appropriate to different types of institutions and programs at different

levels. At the same time, they must make clear just what criteria are being used: e.g., student quality, institutional resources, faculty productivity, the learning environment.

Third, and closely related to the second point, new criteria should be incorporated in assessments of the higher education system: for example, student satisfaction with the educational experience; faculty satisfaction with the academic climate; employer satisfaction with graduates; access and retention; services and benefits to the local community or the state. At the same time, it should be recognized that the importance of these criteria may vary by discipline, educational level, and type of institution.

Fourth, quality assessments should give less emphasis to simply labeling programs and institutions (e.g., "the best," "good," "marginal") and more to pointing the way to improvement. Stronger efforts should be made to identify the special strengths and weaknesses of particular programs and institutions.

Fifth, quality assessment should be dynamic rather than static, taking into consideration not only where a program or institution is now but also where it has come from and where it has the potential to go in the future. Related to this point, institutional officials, aided by accrediting agencies, and by the states, have a responsibility to develop viable implementation plans to assist this kind of long-term institutional development.

Sixth, more attention should be paid to the "value-added" concept of higher education. To give an institution or a program high marks for the resources it is able to attract, without regard for what it does with those resources, is surely to overlook the whole purpose of education at any level: to bring about certain desired changes in students. Before we can judge how well an institution does with and by its students, we must know what the students were like at college entry. The input-environment-outcome model is a conceptual tool whereby the characteristics of entering students can be taken into account to arrive at an assessment of the impact of the college experience itself. Such an approach is especially necessary in a period when the twin doctrines of entitlement and equal educational opportunity are espoused as worthy social goals.

Seventh, failure to address the teaching-learning function perhaps represents the greatest weakness of quality assessments of American higher education.

Over a decade ago, Allan Cartter stated the challenge to which we have only begun to respond adequately:

> Diversity can be a costly luxury if it is accompanied by ignorance. Our present system works fairly well because most students, parents, and prospective employers *know* that a bachelor's degree from Harvard, Stanford, Swarthmore, or Reed is ordinarily a better indication of ability and accomplishment than a bachelor's degree from Melrose A&M or Siwash College. Even if no formal studies were ever undertaken, there is always a grapevine at work to supply impressionistic valuations. However, evaluation by rumor and word of mouth is far from satisfactory . . . Just as consumer knowledge and honest advertising are requisite if a competitive economy is to work satisfactorily, so an improved knowledge of opportunities and of quality is desirable if a diverse educational system is to work effectively.
>
> Evaluation of quality in education, at both the undergraduate and graduate levels, is important not only in determining the front-ranking institutions, but also in identifying lower-ranking colleges. Many prospective graduate students would not be suited to an education at Harvard, the Rockefeller Institute, or California Institute of Technology. Other institutions, in view of their educational offerings, level of work, and quality of students, would provide a happier and more productive experience. Universities, through their selection procedures, and students, though their natural proclivities, tend to sort themselves out into congenial environments. (Cartter 1966, p. 3).

In the expansionist era of the 1960s, the nation could afford to support its many colleges and universities, with their multiplicity of programs, without looking too closely at the contributions they made toward the achievement of desired goals. Now, however, as the college-age population declines in number, as inflation continues to erode financial resources, as the value of higher education comes to be questioned, and as a number of institutions, both public and private, struggle to survive in the changing climate of the 1980s and 1990s, hard decisions will have to be made about what should be retained, what altered, and what eliminated in our current pluralistic system. Thus, the need to define quality in meaningful ways, and to find better means of assessing it, is imperative. Such efforts should be grounded in a commitment to the diversity of the system, an understanding that, if higher education is to serve the needs of a heterogeneous population, diversity is much more than just a "costly luxury."

Bibliography

The ERIC Clearinghouse on Higher Education abstracts and indexes the current literature on higher education for the National Institute of Education's monthly bibliographic journal *Resources in Education* (RIE). Most of these publications are available through the ERIC Document Reproduction Service (EDRS). Ordering number and price for publications cited in this bibliography that are available from EDRS have been included at the end of each citation. Readers who wish to order a publication should write to the ERIC Document Reproduction Service, P.O. Box 190, Arlington, Virginia 22210. When ordering, please specify the document number. Unless otherwise noted, documents are available in both microfiche (MF) and paper copy (PC).

Adams, Arvil V. and Krislov, Joseph. "Evaluating the Quality of American Universities: A New Approach." *Research in Higher Education* 8 (1978): 97-109.

American Education Council, in "Survey of 37 Academic Fields and 130 Schools in U.S., Finds Graduate Facilities Have Improved Substantially Between '64 and '69 at Most Schools." *New York Times*, January 3, 1971, p. 54.

Anderson, Kay J. *Regional Accreditation Standards*. Research Reports, vol. 2. Washington, D.C.: Council on Postsecondary Accreditation, 1978.

Anderson, Richard E. *Strategic Policy Changes at Private Colleges*. New York: Teachers College Press, 1977.

Andrews, Grover C. *Assessing Nontraditional Education*. Washington, D.C.: Council on Postsecondary Accreditation, 1978. ED 165 577. MF-$0.98; PC-$16.69.

Arlt, Gustave O. *A Survey of Master's Level Education in the United States*. Studies in Higher Education for the State of New Jersey. Trenton: New Jersey Department of Higher Education, 1970.

Astin, Alexander W. "College Preferences of Very Able Students." *College and University* (Spring 1965a): 282-97.

———. "Differential College Effects on the Motivation of Talented Students to Obtain the Ph.D." *Journal of Educational Psychology* 54 (1963): 63-71.

———. "Equal Access to Postsecondary Education: Myth or Reality?" *UCLA Educator* 19 (Spring 1977a): 8-17.

———. *Four Critical Years*. San Francisco: Jossey-Bass, 1977b.

———. "Measuring the Outcomes of Higher Education." In *Evaluating Institutions for Accountability*. New Directions for Institutional Research, No. 1, edited by Howard R. Bowen. San Francisco: Jossey-Bass, 1974. pp. 23-46.

———. "The Methodology of Research on College Impact: Part One." *Sociology of Education* 43 (1970): 223-54.

——. *The Myth of Equal Access to Higher Education.* Atlanta: Southern Education Foundation, 1975.
——. *Predicting Academic Performance in College.* New York: Free Press, 1971.
——. "Student-Oriented Management: A Proposal for Change." In *Evaluating Educational Quality: A Conference Summary.* Washington, D.C.: Council on Postsecondary Accreditation, 1979. pp. 3-18. ED 165 648. MF-$0.98; PC-$4.55.
——. "Traditional Conceptions of Quality in Higher Education: It's Time for a Change." Paper presented at the 1980 Conference of the American Association for Higher Education, March 6, 1980, in Washington, D.C. Duplicated.
——. "Undergraduate Achievement and Institutional Excellence." *Science* 161 (1968): 661-668.
——. *Who Goes Where to College.* Chicago: Science Research Associates, 1965b.
—— and Henson, James W. "New Measures of College Selectivity." *Research in Higher Education* 6 (1977): 1-8.
—— and Holland, John L. *The Distribution of "Wealth" in Higher Education.* Evanston, Ill.: National Merit Scholarship Corporation, 1962.
——, King, Margo R., and Richardson, Gerald T. *The American Freshman: National Norms for Fall 1979.* Los Angeles: Graduate School of Education, University of California, Los Angeles, 1979. ED 181 812. MF-$0.98; PC-$17.51.
—— and Panos, Robert J. *The Educational and Vocational Development of College Students.* Washington, D.C.: American Council on Education, 1969.
—— and ——. "The Evaluation of Educational Programs." In *Educational Measurement,* edited by R.L. Thorndike. Washington, D.C.: American Council on Education, 1971. pp. 773-5.
—— and Solmon, Lewis C. "Measuring Academic Quality: An Interim Report." *Change* 11 (1979): 48-51.
Astin, Helen S. and Bayer, Alan E. "Sex Discrimination in Academe." *Educational Record* 53 (Spring 1972): 101-18.
Baird, Leonard. *The Elite Schools: A Profile of Prestigious Independent Schools.* Lexington, Mass.: D.C. Heath, 1977.
Barak, Robert J. *A Survey of State-Level Academic Program Review Policies and Procedures for Higher Education.* Des Moines: Iowa Board of Regents, 1975. ED 107 163. MF-$0.98; PC-$6.37.
—— and Berdahl, Robert O. *State-Level Academic Review in Higher Education.* Denver: Education Commission of the States, 1978. ED 159 638. MF-$0.98; PC-$10.87.
Barron's Profiles of American Colleges. 11th edition. Woodbury, N.Y.: Barron Education Services, 1978.
Bayer, Alan E. *Teaching Faculty in Academe, 1972-73.* Washington, D.C.: American Council on Education, 1973.
—— and Dutton, Jeffrey E. "Career Age and Research-Professional Activities of Academic Scientists." *Journal of Higher Education* 48 (1977): 259-82.
—— and Folger, John K. "Some Correlates of a Citation Measure of Productivity in Science." *Sociology of Education* 39 (1966): 381-90.

Berdahl, Robert O. *Statewide Coordination of Higher Education.* Washington, D.C.: American Council on Education, 1971.

Bess, James L. "Ranking Academic Departments: Some Cautions and Recommendations." *Journal of Higher Education* 42 (1971): 721-27.

Beyer, Janice M. and Snipper, Reuben. "Objective versus Subjective Indicators of Quality." *Sociology of Education* 47 (1974): 541-57.

Blackburn, Robert T. and Gerber, Wayne J. "Expert Rating on Academic Leadership as a Measure of Institutional Quality." *Sociology of Education* 47 (1974): 535-40.

—— and Lingenfelter, Paul E. *Assessing Quality in Doctoral Programs: Criteria and Correlates of Excellence.* Ann Arbor: Center for the Study of Higher Education, University of Michigan, 1973. ED 078 728. MF-$0.98; PC-$6.05.

Blau, Peter M. *The Organization of Academic Work.* New York: John Wiley & Sons, 1973.

—— and Margulies, Rebecca Z. "The Reputations of American Professional Schools." *Change* 6 (Winter 1974-75): 42-47.

Boulding, Kenneth E. "Quality versus Equality: The Dilemma of the University." *Daedalus* 101 (1975): 298-301.

Bowen, Frank M. and Glenny, Lyman A. *State Budgeting for Higher Education.* Berkeley: Center for Research and Development in Higher Education, University of California, 1976. ED 159 978. MF-$0.98; PC-$20.51.

Bowen, Howard R., ed. *Evaluating Institutions for Accountability.* New Directions for Institutional Research, No. 1. San Francisco: Jossey-Bass, 1974.

——. "Goals, Outcomes, and Academic Evaluation." In *Evaluating Educational Quality: A Conference Summary,* pp. 19-28. Washington: Council on Postsecondary Accreditation, 1979. ED 165 648. MF-$0.98; PC-$4.55.

——. *Investment in Learning.* San Francisco: Jossey-Bass, 1977.

Bowker, Albert H. "Quality and Quantity in Higher Education." *Journal of the American Statistical Association* (March 1964): 1-15.

Breneman, David W. *Graduate School Adjustments to the "New Depression" in Higher Education.* Technical Report No. 3. Washington, D.C.: National Board of Graduate Education, 1975. ED 101 643. MF-$0.98; PC-$9.37.

Brown, David G. *The Mobile Professors.* Washington, D.C.: American Council on Education, 1967.

Brush, Stephen G. "The Search for Quality in University Research Programmes." *Social Studies of Science* 7 (1977): 395-400.

California State Department of Education. *A Master Plan for Higher Education in California, 1960-1975.* Sacramento: California State Department of Education, 1960. ED 011 193. MF-$0.98; PC-$17.42.

Callan, Patrick M. "The State Role after *Bakke:* Toward Collaborative Action." In *Admitting and Assisting Students after "Bakke,"* New Directions for Higher Education, No. 23, edited by Alexander W. Astin, Bruce Fuller, and Kenneth C. Green. San Francisco: Jossey-Bass, 1978. pp. 53-60.

Cameron, Kim. "Measuring Organizational Quality in Institutions of Higher Education." *Administrative Quarterly* 23 (1978): 604-632.

Carnegie Council on Policy Studies in Higher Education. *A Classification of Institutions of Higher Education.* Berkeley: Carnegie Foundation for the Advancement of Teaching, 1976.

Carpenter, Ray L., and Carpenter, Patricia A. "The Doctorate in Librarianship and an Assessment of Graduate Library Education." *Journal of Education for Librarianship* 11 (1970): 3-45.

Cartter, Allan A. *An Assessment of Quality in Graduate Education.* Washington, D.C.: American Council on Education, 1966.

—— and Ruhter, Wayne E. *The Disappearance of Sex Discrimination in First Job Placement of New PhDs.* Los Angeles: Higher Education Research Institute, 1975. ED 104 255. MF-$0.98; PC-$4.55.

—— and Solmon, Lewis C. "The Cartter Report on the Leading Schools of Education, Law and Business." *Change* 9 (1977): 44-48.

Cass, James and Birnbaum, Max. *Comparative Guide to American Colleges.* 9th ed. New York: Harper & Row, 1979.

Clark, Burton R.; Heist, Paul; McConnell, T.R.; Trow, Martin A.; and Yonge, George. *Students and Colleges: Interaction and Change.* Berkeley: Center for Research and Development in Higher Education, University of California, Berkeley, 1972. ED 069 255. MF-$0.98; PC-$23.83.

Clark, David L. and Guba, Egon G. *Studies of Productivity in Knowledge Production and Utilization by Schools, Colleges, and Departments of Education.* Bloomington: Indiana University, 1976.

Clark, Kenneth E. *America's Psychologists: A Survey of a Growing Profession.* Washington, D.C.: American Psychological Association, 1957.

Clark, Mary Jo. "The Meaning of Quality in Graduate and Professional Education." In *Scholars in the Making: The Development of Professional Students,* edited by Joseph Katz and Rodney T. Hartnett. Cambridge, Mass.: Ballinger, 1976.

——; Hartnett, Rodney T.; and Baird, Leonard L. *Assessing Dimensions of Quality in Doctoral Education: A Technical Report of a National Study in Three Fields.* Princeton, N.J.: Educational Testing Service, 1976. ED 173 144. MF-$0.98; PC-$30.16.

Clark, Wesley C.; Haskins, Jack B.; Alperstein, Gerald; and Lindberg, Richard A. *Journalism Education Representations.* Syracuse, N.Y.: Syracuse University, 1972.

Clemente, Frank and Sturgis, Richard B. "Quality of Department of Doctoral Training and Research Productivity." *Sociology of Education* 47 (1974): 287-99.

Cole, Jonathan R. and Cole, Stephen. "Measuring the Quality of Sociological Research Problems in the Use of the *Science Citation Index.*" *American Sociologist* 6 (1971): 23-29.

—— and Lipton, James A. "The Reputation of American Medical Schools." *Social Forces* 55 (1977): 662-84.

Cole, R.R. and Bowers, T.A. "Research Article Productivity of U.S. Journalism Faculties." *Journalism Quarterly* 50 (1973): 246-54.

Committee of Fifteen. *The Graduate School, Today and Tomorrow.* New York: Fund for the Advancement of Education, 1955.

Conference Board of Associated Research Councils. *Report of the Planning Committee on a Study of the Quality of Research-Doctorate Programs*. Washington, D.C.: 1978.

Council of Graduate Schools in the United States. *The Assessment of Quality in Graduate Education: Summary of a Multidimensional Approach*. Washington, D.C.: n.d.

———. *Changing Patterns in Graduate Education: Proceedings of the Eighteenth Annual Meeting*. Washington, D.C.: 1978. ED 171 200. MF-$0.98; PC-$20.51.

———. *The Master's Degree: A Policy Statement*. Washington, D.C.: 1976. ED 153 588. MF-$0.98; PC-$3.05.

Council on Postsecondary Accreditation and Council of Graduate Schools in the United States. *Accreditation of Graduate Education*. Washington, D.C.: 1978. ED 153 589. MF-$0.98; PC-$3.05.

Cox, W. Miles and Catt, Viola. "Productivity Ratings of Graduate Programs in Psychology Based on Publication in the Journals of the American Psychological Association." *American Psychologist* 32 (1977): 793-813.

Crane, Diana. "The Academic Marketplace Revisited: A Study of Faculty Mobility Using the Cartter Ratings." *American Journal of Sociology* 75 (1970): 953-64.

———. "Scientists at Major and Minor Universities: A Study of Productivity and Recognition." *American Sociological Review* 30 (1965): 699-714.

Creager, John A. *The Use of Publication Citations in Educational Research*. Washington, D.C.: Office of Research, American Council on Education, 1967. ED 030 363. MF-$0.98; PC-$4.55.

Dickey, Frank G. and Miller, Jerry W. *A Current Perspective on Accreditation*. Washington, D.C.: American Association for Higher Education, 1972. ED 068 071. MF-$0.98; PC-$6.05.

Dolan, W. Patrick. *The Ranking Game: The Power of the Academic Elite*. Lincoln: University of Nebraska Press, 1976. ED 129 131. MF-$0.98; PC-$7.87.

Dressel, Paul. *Problems and Principles in the Accreditation of Graduate Education*. Washington, D.C.: Council on Postsecondary Accreditation, 1978. ED 165 568. MF-$0.98; PC-$6.05.

——— and Mayhew, Lewis B. *Higher Education as a Field of Study*. San Francisco: Jossey-Bass, 1975.

Drew, David E. *Science Development: An Evaluation Study*. Washington, D.C.: National Academy of Sciences, 1975. ED 108 967. MF-$0.98; PC-$14.19.

——— and Karpf, Ronald S. *Evaluating Science Departments: A New Index*. Santa Monica, Calif.: Rand Corporation, 1975. ED 120 008. MF-$0.98; PC-$4.55.

Dubé, W.F. "Undergraduate Origins of U.S. Medical Students." *Journal of Medical Education* 49 (1974): 1005-10.

Education Commission of the States. *The States and Graduate Education*. Denver, Col.: 1975a. ED 101 644. MF-$0.98; PC-$4.55.

———. "Survey of the Structure of State Coordinating or Governing Boards and Public Institutional and Multi-Campus Governing Board of Postsecondary Education—as of January 1, 1975." *Higher Education and the States* 4 (1975b): 297-352. ED 105 805. MF-$0.98; PC-$6.05.

Eells, Walter C. "Leading American Graduate Schools." *Association of American Colleges Bulletin* 43 (1957): 563-76.

———. "Leading American Graduate Schools, 1948-1958." *Liberal Education* 46 (March 1960): 16-20.

Elton, Charles F. and Rodgers, Samuel A. "The Departmental Rating Game: Measure of Quantity or Quality?" *Higher Education* 2 (1973): 439-46.

——— and ———. "Physics Department Ratings: Another Evaluation." *Science* 174 (1971): 565-68.

——— and Rose, H. A. "What Are Ratings Rating?" *American Psychologist* 27 (1972): 197-201.

"Evaluating Gradute Programs." *Chronicle of Higher Education*, December 6, 1976, pp. 1, 11.

Ewell, Raymond H. "A Qualified Summary of the American Council's Evaluation Report *An Assessment of Quality in Graduate Education.*" Buffalo: State University of New York at Buffalo, 1966. Duplicated.

Fiedler, Fred E. and Biglan, Anthony. *The First Job After Graduation as a Measure of Departmental Performance.* Technical Report No. 74. Urbana, Ill.: Group Effectiveness Research Laboratory, 1969.

"The 15 Top-Ranked Business Schools in the United States." *MBA* (December 1974): 21-25.

Furniss, W. Todd. "Responding with Quality." In *The Closing System of Academic Employment.* Atlanta: Southern Regional Education Board, 1978. ED 160 032. MF-$0.98; PC-$7.87.

Gallant, Jonathan A. and Prothero, John W. "Weight-watching at the University: The Consequences of Growth." *Science* 175 (1972): 381-88.

Gappa, Judith M. and Uehling, Barbara S. *Women in Academe: Steps to Greater Equality.* AAHE-ERIC/Higher Education Research Report No. 1. Washington, D.C.: American Association for Higher Education, 1979. ED 169 873. MF-$0.98; PC-$7.87.

Giordano, Anthony B. "Assessing the Quality of Graduate Programs." *Engineering Education* 67 (1976): 253-56.

Glazer, Nathan. "The Schools of the Minor Professions." *Minerva* 12 (1974): 346-64.

Glenn, Norval D. and Villemez, Wayne. "The Productivity of Sociologists at 45 American Universities." *American Sociologist* 5 (1970): 224-52.

Glenny, Lyman A.; Shea, John R.; Ruyle, Janet H.; and Freschi, Kathryn H. *Presidents Confront Reality: From Edifice Complex to University Without Walls.* A Report for the Carnegie Council on Policy Studies in Higher Education. San Francisco: Jossey-Bass, 1976.

Goldstein, E. "Effect of Same-Sex and Cross-Sex Role Models on the Subsequent Academic Productivity of Scholars: *American Psychologist* 34 (May 1979): 407-10.

Gourman, Jack. *The Gourman Report: A Rating of Graduate and Professional Programs in American and International Universities.* Los Angeles: National Education Standards, 1977a.

——. *The Gourman Report: A Rating of Undergraduate Programs in American and International Universities.* Los Angeles: National Education Standards, 1977b.

——. *The Gourman Report: Ratings of American Colleges.* Phoenix, Ariz.: Continuing Research Institute, 1967.

Gregg, Russell T. and Sims, Paul D. "Quality of Faculties and Programs of Graduate Departments of Educational Administration." *Educational Administration Quarterly* 8 (1972): 67-92.

Guba, Egon G. and Clark, David L. "Levels of R&D Productivity in Schools of Education." *Educational Researcher* 7 (1978): 3-9.

Gurman, Alan S. "Graduate Schools Origins of Authors in the *Journal of Consulting and Clinical Psychology.*" *Journal of Consulting and Clinical Psychology* 39 (1972): 491-94.

Hagstrom, Warren O. "Inputs, Outputs and the Prestige of University Science Departments." *Sociology of Education* 44 (1971): 375-97.

Halstead, D. Kent. *Statewide Planning in Higher Education.* Washington, D.C.: Government Printing Office, 1974. ED 096 914. MF-$1.51; PC-$58.66.

Harcleroad, Fred F. *Educational Auditing and Accountability.* Washington. D.C.: Council on Postsecondary Accreditation, 1976. ED 131 765. MF-$0.98; PC-$4.55.

——. *Accreditation: History, Process, and Problems.* AAHE-ERIC/Higher Education Research Report No. 6, Washington, D.C.: American Association for Higher Education, 1980 (forthcoming).

—— and Dickey, Frank G. *Educational Auditing and Voluntary Institutional Accrediting.* Washington, D.C.: American Association for Higher Education, 1975. ED 102 919. MF-$0.98; PC-$4.55.

Hargens, Lowell L. "Patterns of Mobility of New Ph.D.'s Among American Academic Institutions." *Sociology of Education* 42 (1969): 18-37.

Harris, John. "Critical Characteristics of an Accreditable Institution, Basic Purposes of Accreditation, and Nontraditional Forms of Most Concern." In *Research Reports,* vol. 2 of the Project to Develop Evaluative Criteria and Procedures for the Accreditation of Nontraditional Education. Washington, D.C.: Council on Postsecondary Accreditation, September 1978, pp. 61-105.

Harris, Laurie and Jefferson, Joseph W. "How Researchers Rank Them." *MBA* (December 1974): 24.

Hartnett, Rodney T. and Centra, John A. "The Effects of Academic Departments on Student Learning." *Journal of Higher Education* 48 (1977): 491-507.

——; Clark, Mary Jo; and Baird, Leonard L. "Reputational Ratings of Doctoral Programs." *Science* 199 (1978): 1310-14.

Harvey, James. *The Stuuent in Graduate School.* Washington, D.C.: American Association for Higher Education, 1972. ED 057 258. MF-$0.98; PC-$6.05.

Hawes, Gene R. *Hawes Comprehensive Guide to Colleges.* New York: New American Library, 1978.

Heyns, Roger W., ed. *Leadership for Higher Education.* Washington, D.C.: American Council on Education, 1977.

Higgins, Arthur S. "The Rating of Selected Fields of Doctoral Study in the Graduate School of Education: An Opinion Survey." Ph.D. dissertation, Teachers College, Columbia University, 1968.

Hill, Joan A.; Lutterbie, P.H.; and Stafford, J.H. "Systemwide Academic Review: The Florida Plan." Paper presented at the annual meeting of the American Association for Higher Education, Washington, D.C., April 1979. Duplicated.

"How Professors Rated Faculty in 19 Fields." *Chronicle of Higher Education*, January 15, 1979, pp. 6-7.

Hughes, Raymond M. "Report of the Committee on Graduate Instruction." *Educational Record* 15 (1934): 192-234.

———. *A Study of the Graduate Schools in America*. Oxford, Ohio: Miami University Press, 1925.

Hurlbert, Beverly M. "Status and Exchange in the Profession of Anthropology." *American Anthropologist* 78 (1976): 272-84.

Jacobson, Robert L. "It Does, It Doesn't Matter Where You Go to College." *Chronicle of Higher Education*, November 8, 1976, pp. 1, 12.

———. "Ratings of Schools on a Continual Basis Proposed." *Chronicle of Higher Education*, May 24, 1971, p. 4.

Jencks, Christohper and Riesman, David. *The Academic Revolution*. Garden City, N.Y.: Doubleday, 1968.

Johnson, Richard R. "Innovation and Influence among American Colleges and Universities." Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1979. Duplicated. ED 174 147. MF-$0.98; PC-$4.55.

———. "Leadership among American Colleges." *Change* (November 1978a): 50-51.

———. "The Myth of Leadership among Colleges and Universities." *The North Central Association Quarterly* 54 (Summer 1979): 25-32.

Jordan, Robert T. "Library Characteristics of Colleges Ranking High in Academic Excellence." *College and Research Libraries* 24 (1963): 369-76.

Keeton, Morris. "An Approach to the Theory of Quality Assurance." Remarks to the Conference on Quality Control in Nontraditional Higher Education, Antioch College, 1974. ED 098 868. MF-$0.98; PC-$3.05.

Kelso, Charles D. "How Does Your Law School Measure Up?" *Student Lawyer* 4 (1975): 20-24.

Keniston, Hayward. *Graduate Study and Research in the Arts and Sciences at the University of Pennsylvania*. Philadelphia: University of Pennsylvania Press, 1959.

Kerr, Clark. "Comparative Effectiveness of Systems: Unknown and Unknowable." In *Twelve Systems of Higher Education: Six Decisive Issues*, by Clark Kerr, Burton Clark, Brian MacArthur, and Howard Bowen. New York: International Council for Educational Development, 1978, pp. 157-77.

———. *The Uses of the University*. Cambridge, Mass.: Harvard University Press, 1963.

Knudsen, Dean D. and Vaughan, Ted R. "Quality in Graduate Education: A Re-evaluation of the Rankings of Sociology Departments in the Cartter Report." *American Sociologist* 4 (1969): 12-19.

Krause, Ervin D. and Krause, Loretta. "The Colleges that Produce our Best Scientists: A Study of the Academic Training Grounds of a Large Group of Distinguished American Scientists." *Science Education* 54 (1970): 133-40.

Lakoff, Sanford. "Accountability and the Research Universities." In *The State of Academic Science*, Vol. 2, *Background Papers*, edited by Bruce L.R. Smith and Joseph J. Karlesky. New Rochelle, N.Y.: Change Magazine Press, 1978.

Lewis, Alfred J. "1973 Statistical Survey of Law School Libraries and Librarians." *Law Library Journal* 67 (1974): 283-99.

Lewis, Lionel S. "On Subjective and Objective Rankings of Sociology Departments." *American Sociologist* 3 (1968): 129-31.

Leys, Wayne. "The Terminal Master's Degree." *Harvard Educational Review* 26 (1956): 233-40.

Longanecker, David A. "Assessing Quality in Graduate Education: A Multivariate Approach." Ph.D. dissertation, School of Education, Stanford University, 1978.

———. "The Assessment of Quality in Graduate Education: Summary of A Multidimensional Approach." Washington, D.C.: Council of Graduate Schools in the United States, 1979.

Magoun, H.W. "The Cartter Report on Quality in Graduate Education." *Journal of Higher Education* 37 (1966): 481-92.

Margolis, J. "Citation Indexing and Evaluation of Scientific Papers." *Science* 155 (1967): 1213-19.

Margulies, Rebecca Z. and Blau, Peter. "America's Leading Professional Schools." *Change* 5 (1973): 21-27.

Martorana, S.V. and Hollis, E.V. *State Boards Responsible for Higher Education*. Washington, D.C.: Government Printing Office, 1960.

——— and Kuhns, Eileen. *Managing Academic Change*. San Francisco: Jossey-Bass, 1975.

Marty, Martin. "Rating Welch's Ratings." *Christian Century* 89 (1972): 122-23.

Mayhew, Lewis B. and Ford, Patrick J. *Reform in Graduate and Professional Education*. San Francisco: Jossey-Bass, 1974.

Meyerson, Martin. "After a Decade of Levelers in Higher Education: Reinforcing Quality While Maintaining Mass Education." *Daedalus* 101 (1975): 304-21.

Mincer, Jacob. "The Determination of Labor Incomes: A Survey with Special Reference to the Human Capital Approach." *Journal of Economic Literature* 8 (1970): 1-26.

Mingle, James R. "Influencing Academic Outcomes: The Power and Impact of Statewide Program Review." In *The Closing System of Academic Employment*. Atlanta: Southern Regional Education Board, 1978. ED 160 032. MF-$0.98; PC-$7.87.

Morgan, David R.; Kearney, Richard C.; and Regens, James L. "Assessing Quality among Graduate Institutions of Higher Education in the United States." *Social Science Quarterly* 57 (1976): 671-79.

Mortimer, Kenneth P. *Accountability in Higher Education.* Washington, D.C.: American Association for Higher Education, 1972. ED 058 465. MF-$0.98; PC-$6.05.

"Most Selective Institutions of Higher Education." *HERI Quarterly* 1 (1978): 3.

Muffo, J. "Relationships Between Colleague Rating and Faculty Compensation." *Research in Higher Education* 10 (1979): 25-35.

Munson, C. Eric and Nelson, Pat. "Measuring the Quality of Professional Schools." *UCLA Educator* 19 (1977): 41-52.

Myers, C. Roger. "Journal Citations and Scientific Eminence in Contemporary Psychology." *American Psychologist* 25 (1970): 1041-48.

National Board on Graduate Education. *Graduate Education: Purposes, Problems, and Potential.* Washington, D.C.: 1972. ED 070 413. MF-$0.98; PC-$4.55.

——. *Science Development, University Development, and the Federal Government.* Washington, D.C.: 1975. ED 108 968. MF-$0.98; PC-$6.05.

National Commission on Accreditation. *Study of Accreditation in Selected Health Programs.* Washington, D.C.: 1972. ED 068 628. MF-$0.98; PC-$6.05.

National Center for Education Statistics. *Projections of Education Statistics to 1988-89.* Washington, D.C.: Government Printing Office, 1980.

National Science Foundation, National Science Board. *Graduate Education: Parameters for Public Policy.* Washington, D.C.: 1969. ED 041 560. MF-$0.98; PC-$14.19.

Newman, Frank, chairman, Task Force on Higher Education. *Report on Higher Education* [The First Newman Report]. Submitted to the Office of Education, Department of Health, Education, and Welfare. Washington, D.C.: Government Printing Office, 1971. ED 049 718. MF-$0.98; PC-$10.87.

Nichols, J.H. "The Welch Report and the Seminaries." *Christian Century* 89 (1972): 122-23.

Nichols, Robert C. *College Preferences of Eleventh Grade Students.* National Merit Scholarship Corporation Research Reports, vol. 2, no. 9. Evanston, Ill.: 1966.

Olivas, Michael A. *The Dilemma of Access.* Washington, D.C.: Howard University Press, 1979.

Oromaner, Mark J. "A Note on Analytical Properties and Prestige of Sociological Departments." *American Sociologist* 5 (1970): 240-44.

Packer, James S. and Murdoch, Wallace P. "Publication of Scientific Papers in the Journals of the Entomological Society of America: An 11-Year Review." *Bulletin of the Entomological Society of America* 20 (1974): 249-53.

Petersen, Dorothy G. "Accrediting Standards and Guidelines: A Profile." *Educational Record* 59 (1978): 305-13.

Petrowski, William R.; Brown, Evan L.; and Duffy, John A. "National Universities and the ACE Ratings." *Journal of Higher Education* 44 (1973): 495-513.

Pirsig, Robert M. *Zen and the Art of Motorcycle Maintenance.* New York: William Morrow, 1974.

"The Popular Vote: Rankings of Top Schools." *Juris Doctor* 6 (1976): 17-18, 21.

Potts, Edwin. *PDS Goals and Climate Manual.* Washington, D.C.: Council for the Advancement of Small Colleges, forthcoming.

Powel, John M., Jr. and Lamson, Robert D. *An Annotated Bibliography of Literature Relating to the Costs and Benefits of Graduate Education.* Washington, D.C.: Council of Graduate Schools in the United States, 1972. ED 069 200. MF-$0.98; PC-$6.05.

Psacharopoulos, George. *Earnings and Education in OECD Countries.* Paris Organisation for Economic Cooperation and Development [1975]. ED 121 959. MF-$0.98; PC-$14.19.

Regents Commission on Doctoral Education. *Meeting the Needs of Doctoral Education in New York State* [The Fleming Commission Report]. Albany: New York State Board of Regents, 1973. ED 083 936. MF-$0.98; PC-$7.87

Regents of the University of the State of New York. *Meeting the Needs of Doctoral Education.* Albany: New York State Education Department, 1973. ED 083 897. MF-$0.98; PC-$4.55.

Roose, Kenneth D. and Andersen, Charles J. *A Rating of Graduate Programs.* Washington, D.C.: American Council on Education, 1970.

Selden, William K. *Accreditation.* New York: Harper & Bros., 1960.

———. *Accreditation and the Public Interest.* Washington, D.C.: Council on Postsecondary Accreditation, 1976. ED 131 766. MF-$0.98 PC-$4.55.

——— and Porter, Harry V. *Accreditation: Its Purposes and Uses* Washington, D.C.: Council on Postsecondary Accreditation, 1977. ED 142 115. MF-$0.98; PC-$4.55.

Shichor, David. "Prestige of Sociology Departments and the Placing of New Ph.D.'s." *American Sociologist* 5 (1970): 157-60.

Shulman, Carol H. *Compliance with Federal Regulations: At What Cost?* AAHE-ERIC/Higher Education Research Report No. 6. Washington, D.C.: American Association for Higher Education, 1978. ED 165 552. MF-$0.98; PC-$6.05.

———. *Old Expectations, New Realities: The Academic Profession Revisited.* AAHE-ERIC/Higher Education Research Report No. 2. Washington, D.C.: American Association for Higher Education, 1979. ED 169 874. MF-$0.98; PC-$6.05.

Siebring, B. Richard. "The American Council on Education Rankings of Quality in Graduate Education and Membership in the National Academy of Science." *Science Education* 53 (1969): 73-77.

Smith, Richard and Fiedler, Fred E. "The Measurement of Scholarly Work: A Critical Review of the Literature." *Educational Record* 52 (1971): 225-32.

Solmon, Lewis C. "The Definition of College Quality and its Impact on Earnings." *Explorations in Economic Research* 2 (1975): 537-87.

———. *Male and Female Graduate Students: The Question of Equal Opportunity.* New York: Praeger, 1976. ED 153 566. MF-$0.98; PC-$12.69.

———. "Schooling and Subsequent Success." In *Does College Matter?*, edited by Lewis C. Solmon and Paul J. Taubman. New York: Academic Press, 1973. pp. 13-34.

—— and Ochsner, Nancy I. *New Findings on the Effects of College.* Current Issues in Higher Education. Washington, D.C.: American Association for Higher Education, 1978.

Solmon, W.E. and Walters, A.T. "The Relationship Between Productivity and Prestige of Graduate Sociology Departments: Fact or Artifact?" *American Sociologist* 10 (1976): 229-36.

Somit, Alfred and Tanenhaus, Joseph. *American Political Science: A Profile of a Discipline.* New York: Atherton Press, 1964.

State of New York, Chemistry Program Evaluation Committee. "Report to the Commissioner of Education of the State of New York." Albany, n.d. Mimeographed.

Sturdivant, F.D. and Adler, R.D. "Executive Origins: Still a Gray Flannel World?" *Harvard Business Review* 54 (1976): 125-32.

Thrash, Patricia. "Accreditation: A Perspective." *Journal of Higher Education* 50 (1979): 115-21.

Tidball, M. Elizabeth. "Perspective on Academic Women and Affirmative Action." *Educational Record* 54 (1973): 130-35.

—— and Kistiakowski, V. "Baccalaureate Origins of American Scientists and Scholars." *Science* 193 (1976): 646-52.

Toombs, William E. *Productivity: The Burden of Success.* Washington, D.C.: American Association for Higher Education, 1975. ED 076 174. MF-$0.98; PC-$6.05.

"The Top 15: The MBA Survey of Graduate Business Schools." *MBA* (December 1975): 33-35.

Trivett, David A. *Accreditation and Institutional Eligibility.* AAHE-ERIC/Higher Education Research Report No. 9. Washington, D.C.: American Association for Higher Education, 1976. ED 132 919. MF-$0.98; PC-$9.37.

Troutt, W.E. "Regional Accreditation Evaluative Criteria and Quality Assurance." *Journal of Higher Education* 50 (1979): 199-210.

Trow, Martin. "The Public and Private Lives of Higher Education." *Daedalus* 101 (1975): 113-27.

Tyler, Forrest B. "Knowledgeable Respondents: Private Club or Public Service?" *American Psychologist* 27 (1972): 191-96.

U.S. Department of Health, Education and Welfare. Advisory Committee on Accreditation and Institutional Eligibility. *Annual Report, 1977.* Washington, D.C.: Government Printing Office, 1977.

Walsh, W. Bruce. *Theories of Person-Environment Interaction: Implications for the College Student.* Iowa City: American College Testing Program, 1973.

——; Feeney, Dorothy; and Resnick, Harvey. "Graduate School Origins of *Journal of Counseling Psychology* Authors." *Journal of Counseling Psychology* 16 (1969): 373-76.

Wanderer, Jules J. "Academic Origins of Contributors to the *American Sociological Review,* 1955-1965." *American Sociologist* 1 (1966): 241-43.

Webster, David. "A Brief History and Critique of Published Ratings of Graduate School Departments." Seminar paper, Graduate School of Education, University of California at Los Angeles, 1977.

——. "How to Assess Quality in Master's Programs." In *Contributed Papers*, National Conference on the Assessment of Quality in Master's Programs, University of Maryland, 1979. pp. 71-80.

Welch, Claude. *Graduate Education in Religion: A Critical Appraisal.* Missoula: University of Montana Press, 1971.

"Who Got U.S. Funds: The Top 100 Colleges." *Chronicle of Higher Education*, September 3, 1974, p. 8.

Wispe, Lauren G. "The Bigger the Better: Productivity, Size, and Turnover in a sample of Psychology Departments." *American Psychologist* 24 (1969): 662-68.

Wolfle, Dael. "Forces Affecting the Research Role of Universities." In *The State of Academic Science*, vol. 2, *Background Papers*, edited by Bruce L.R. Smith and Joseph J. Karlesky. New Rochelle, N.Y.: Change Magazine Press, 1978.

Wong, S. Godwin. "Pecking Orders: Uses and Limitations." Paper presented at the 17th Annual Forum of the Association for Institutional Research. Montreal, Quebec, May 1977. ED 146 861. MF-$0.98; PC-$3.05.

Young, Kenneth. "Accreditation and Graduate Education." In Proceedings of the Sixteenth Annual Meeting of the Council of Graduate Schools in the United States. Washington, D.C.: 1976a. pp. 133-8.

——. "Evaluating Institutional Effectiveness." *Educational Record* 57 (1976b): 45-52.

AAHE-ERIC Research Reports

Ten monographs in the AAHE-ERIC Research Report series are published each year, available individually or by subscription. Single copies are $3 for AAHE members, $4 each for nonmembers. Orders for 25 or more are eligible for bulk discounts. Add $1 postage/handling fee for all orders under $15. Orders under $15 must be prepaid.

Subscription to ten issues (beginning with date of subscription) is $25 for AAHE members, $35 for nonmembers. Order from Publications Department, American Association for Higher Education, One Dupont Circle, Suite 780, Washington, D.C. 20036; 202/293-6440. Write or phone for a complete list of Research Reports and other AAHE publications.

1979 Research Reports

1. Women in Academe: Steps to Greater Equality
   Judith Gappa and Barbara Uehling
2. Old Expectations, New Realities:
   The Academic Profession Revisited
   Carol Herrnstadt Shulman
3. Budgeting in Higher Education
   J. Kent Caruthers and Melvin Orwig
4. The Three "R's" of the Eighties: Reduction,
   Retrenchment, and Reallocation
   Kenneth P. Mortimer and Michael L. Tierney
5. Occupational Programs in Four-Year Colleges:
   Trends and Issues
   Dale F. Campbell and Andrew S. Korim
6. Evaluation and Development of Administrators
   Robert C. Nordval
7. Academic Advising: Getting Us Through the Eighties
   Thomas J. Grites
8. Professional Evaluation in the Eighties:
   Challenges and Response
   Glenn F. Nyre and Kathryn C. Reilly
9. Adult Baccalaureate Programs
   Marilou Denbo Eldred and Catherine Marienau
10. Survival Through Interdependence: Assessing the Cost and
    Benefits of Interinstitutional Cooperation
    Lewis D. Patterson

1980 Research Reports

1. Federal Influence on Higher Education Curricula
   William V. Mayville
2. Program Evaluation
   Charles E. Feasley
3. Liberal Education in Transition
   Clifton E. Conrad and Jean C. Wyer
4. Adult Development: Implications for Higher Education
   Rita Preszler Weathersby and Jill Mattuck Tarule
5. A Question of Quality: The Higher Education Rating Game
   Judith K. Lawrence and Kenneth C. Green